MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2870
Measure Title: Overuse of Opioid Containing Medications for Primary Headache Disorders
Measure Steward: American Academy of Neurology
Brief Description of Measure: Percentage of patients aged 12 years and older diagnosed with primary headache disorder, and taking an opioid containing medication who were assessed for opioid containing medication overuse within the 12-month measurement period, and treated or referred for treatment if identified as overusing opioid containing medication.
Developer Rationale: Triptans and ergots are considered first line acute treatments for migraine, not opioids or barbiturates by the US Headache Consortium Guideline. The use of barbituates or opioids increases the risk of chronic daily headache and drug induced hyperalgesia. In one study, any use of barbituates and opiates was associated with increased risk of transformed migraine after adjusting for covariates, while triptans were not. In a sample of 5, 796 people with headache, 4,076 (70.3%) were opioid nonusers, 798 (13.8%) were previous users, and 922(15.9%) were current opioid users.

Rates of headache-related health-care resource utilization were higher for all opioid-use groups for emergency department/urgent care, primary care, and specialty care visits compared to nonusers. Opioid use for migraine is associated with more severe headache-related disability, symptomology, comorbidities (depression, axiety, and cardiovascular disease and events), and greater health-care resource utilization for headache.
- Average headache frequency in days per month was higher among current opioid users
- Previous and current opioid users "had significantly higher average Migraine Disability Assessment (MIDAS) scores when compared to nonusers.
- Comorbidities tended to occur at higher rates among all opioid-use groups compared to nonusers, with depression significantly more common among opioid-use groups compared to nonusers.
- Health-care resource use was higher among previous and current opioid user

Numerator Statement: Patients assessed for opioid containing medication overuse within the 12-month measurement period and treated or referred for treatment if identified as overusing opioid containing medication which is defined as: Using opioid containing medication for greater than or equal to 10 days per month for more than 3 months.
Denominator Statement: All patients aged 12 years and older diagnosed with a primary headache disorder* and taking opioid containing medication.

*Define Primary Headache: A headache that is not caused by another disease or medical condition. For the purpose of this measure this includes the following types of headache:
Migraine - Migraine without aura, migraine with aura, childhood periodic syndromes that are commonly precursors of migraine, retinal migraine, complications of migraine, probable migraine.
Tension-Type Headache (TTH) - Infrequent episodic TTH, frequent episodic TTH headache, chronic TTH, probable TTH
Cluster Headache (CH) and Other Trigeminal Autonomic Cephalgias: Cluster headache, paraxysmal hemicrania, short-lasting unilateral neuralgia from headache attacks with conjuctival injection and tearing (SUNCT), probably trigeminal autonomic cephalgia
Other Primary Headaches: Primary stabbing headache, primary cough headache, primary exertional headache, primary

headache associated with sexual activity, hypnic headache, primary thunderclap headache, hemicrania continua, new daily-persistent headache.
Denominator Exclusions: No Exclusions.

Medical exceptions for not assessing, treating, or referring patient for treatment of opioid medication overuse include:
Patient already assessed and treated for opioid use disorder within the last year.
Patient has a documented failure of non-opioid options and is not identified as overusing opioid containing medication.
Patient has contraindications to all other medications for primary headache.

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry
Level of Analysis: Clinician : Individual

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☐ **Yes** ☒ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☒ **No**
- **Evidence graded?** ☐ **Yes** ☒ **No**

Evidence Summary

The developer cites four clinical practice guidelines that recommend against the use of opioids in primary headache disorders and one clinical guideline that suggests referral to a neurologist/specialist. One of the references listed as a clinical guideline is questionable regarding classification as a guideline and appears to be more of a summary of the evidence; despite this it does appear to support the harms of overutilization of opioids. An open area of question is if the guideline evidence supports the remaining measure concepts: assessed for overuse and treatment or referral.

This measure focus seems to be on the interventions of 1) assessing for overuse of opioid medications in a typical headache patient population and 2) for those patients identified as being over-users of opioid medications there should be evidence of treatment for that overuse or a documentation of referral for treatment. The evidence submitted substantiates the harms and clinical guideline recommendations against opioid use, but does not support the processes (interventions) of referral and treatment

Exception to evidence
N/A

Questions for the Committee:

- *Does the Committee agree there was no evidence submitted to support the full measure conceptual model? Assessment – identification – referral/treatment? Would the Committee rate the evidence as low? Or insufficient?*
- *Did the developer clearly state the relationship of the measure to patient outcomes?*
- *Is there evidence available to support the identification of opioid over-utilizers and the impact of*

referral/treatment on the intended outcomes?

- *For possible exception to the evidence criterion:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
  - *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
  - *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empirical evidence?*

Guidance from the Evidence Algorithm

Process measure/systematic review (Box 3) No (evidence is about something other than what is being measured)→Empirical Evidence submitted but without systematic review and grading of the evidence (Box 7) No→performance measures of related health outcome or intermediate outcome or process (Box 10) Yes →Insufficient (premise of the measure may not be incorrect, the issue is that evidence was not provided that supports the interventions leading to decrease in outcome (reduced use of opioids in typical headaches).

Preliminary rating for evidence: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

Preliminary rating for evidence: ☐ **Pass** ☒ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

Information on performance of measure, as specified, was not provided in the measure submission form. Data cited regarding the overuse of opioid use in migraine sufferers indicated that in one study, 13.8% of patients had evidence of previous use and 15.9% current use. The developers did provide 4 citations regarding the treatment and management of migraines/headaches, but a summary was not provided to assess impact, quality gap or opportunity for improvement. Prevalence and cost data support potential impact of improved management of migraine/headaches.

Disparities

Only general prevalence data was provided: In US population studies, the prevalence of migraine is approximately 18% in women and 6% in men. The prevalence of migraine in children is 7.7%.

Questions for the Committee:

- *In evaluating performance gap, the Committee should consider the demonstration of quality problems and opportunity for improvement. The developers provide some data on the prevalence of migraines and the use of opioids in the population, but very little additional data. Is there additional data that demonstrate opportunity for improvement or quality problems specific to the assessment, referral and treatment of overuse of opioids?*
- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **Insufficient evidence

**This a very important issue, but the data in support of the measure are very limited. As pointed out by Latana, the specific criterion has a somewhat arbitrary nature which will doubtless miss important over-prescriptions or inappropriate prescriptions. How does one deal with self-medication with now commonly available street drugs - opiates and otherwise? As a trial this may be worthy of support to promote further research but I would like to see some of the conceptual issues resolved or better supported by data before going there. The referral to treatment criterion, while measurable seems a very mushy test of overuse - especially with the potential for patients to find street drug substitutes or other analgesics.

**insufficient

1b. Performance Gap

Comments: **No data on performance

**No data presented - There is doubtless excessive opiate prescription for many things, including primary headache. My bias is to focus on the incidence of appropriate treatments of primary headache rather than on the overuse of inappropriate treatments with various potential adverse consequences.

**insufficient

1c. High Priority (previously referred to as High Impact)

Comments: **NA

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic clinical data, EHR

Specifications: Patients assessed for opioid containing medication overuse within the 12-month measurement period and treated or referred for treatment if identified as overusing opioid containing medication which is defined as: Using opioid containing medication for greater than or equal to 10 days per month for more than 3 months. The denominator is All patients aged 12 years and older diagnosed with a primary headache disorder* and taking opioid containing medication. There are no exclusions, but there are medical exceptions that could be found in the EHR. The measure is not risk adjusted.

This is a new eMeasure – HQMF specification are included. See eMeasure Technical Review below

Questions for the Committee :

- *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor review:

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No** **eMeasure Trial Approval Technical Review has found that:** • The submitted eMeasure specification captures the data elements and measure logic needed for automated measure calculation
Documentation of HQMF or QDM	All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM;

limitations	
Value Sets	Value sets used in the submitted eMeasure are published in the NLM Value Set Authority Center.
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously
Feasibility Testing	The feasibility analysis submitted by the measure developer meets the requirements to be considered for eMeasure Trial Approval.

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Method(s) of reliability testing This is an emeasure being considered for Trial approval, as such reliability and validity testing results are not needed for submission. The developer has indicated they are partnering with Minnesota Community Measurement to test this measure at the individual clinician level. See notes about emeasure technical review above.

The Committee will not be asked to vote on Reliability

Questions for the Committee:

- *The developer provides a brief summary of testing plans, do you have specific questions on the planned method of reliability testing?*

Preliminary rating for reliability: ☐ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☐ **Yes** ☐ **Somewhat** ☒ **No**
Specification not completely consistent with evidence

The specifications for this measure report if patients are assessed for overuse of opioids, and then if referral or treatment occurred. The evidence provided supports that opioids may be commonly overused in the headache population and should not be; but evidence was not provided on the relationship between assessment, and treatment or referral to the actual outcome – which is assumed to be better headache management.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

Validity will be tested by both Minnesota Community measurement and in the future through the AAN's Axon

Registry. The Committee will not be asked to vote on Validity for this emeasure under consideration for trial approval.

Questions for the Committee:

- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

Exclusions and the burden of exclusions included in the measures are evaluated not only as the e-measures are developed but on an ongoing basis. The AAN has a system in place to garner user feedback from those members participating in the AAN's Axon Registry, burden of exclusions is an issue the AAN will continue to assess participants of the registry.

The currently specified exclusions are:

Medical exceptions for not assessing, treating, or referring patient for treatment of opioid medication overuse include:

Patient already assessed and treated for opioid use disorder within the last year.

Patient has a documented failure of non-opioid options and is not identified as overusing opioid containing medication.

Patient has contraindications to all other medications for primary headache.

Questions for the Committee:

- *Are there any exclusions you would expect to see in this measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

N/A

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

N/A

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **No

**Not voting here. I see major difficulties in establishing a good clean validity measure.

2a2. Reliability Testing

Comments: **NA

**Not voting here.

**Validity is limited by variability of patients and settings.

Re. Patients

-- comorbidities may be significant factor in opioid decision

--acute versus chronic h/o headache may be different populations

Re. Settings

--Differences underlying opioid decision between ER versus Outpatient Care versus Inpatient Care may require additional supporting evidence

2b2. Validity Testing

Comments: **NA

**As mentioned above, data on alternative treatments (including self-medications, and "alternative" therapies) are a problem that needs to be address - though I suspect with considerable difficulty

**variability of patients and settings limits broad application of this measure until more is known about homogeneity/heterogeneity of headache opioid-treatment population.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **NA

**Not voting here.

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

This measure is proposed as an eMeasure for trial use, as such, it underwent an feasibility assessment by an eMeasure technical advisor. The developer also provided information from testing conducted by a contractor (Lantana) regarding feasibility:

- The Lantana team shared potential challenges in the measure related to data standards:
  - Non-use codes for treatment—Currently this concept is not captured discretely in Behavioral Health Information Technology systems. According AAN, neurologists do not currently utilize any standard codes. Candidate for revision is Referral for Opioid Overuse Treatment SNOMEDCT Value Set; its associated OID is 2.16.840.1.113762.1.4.1111.90.
  - Failed treatment—This procedure is not consistently documented or captured in EHRs. Candidate for revision is Treatment Failure SNOMEDCT Value Set; its associated OID is 2.16.840.1.113762.1.4.1111.93.
  - Opioid assessments—There is no standard form or question that captures data related to whether an opioid assessment was conducted. Candidate for revision is Opioid Overuse Assessment SNOMEDCT Value Set; its associated OID is 2.16.840.1.113762.1.4.1111.89.
  - Opioid overuse—The current definition of this concept captures only overuse fitting the precise description. Example: patient takes opioids three times a day, every day for 2 months, would not be considered as overusing.
  - Calculating cumulative opioid use—The QDM is limited in its ability to calculate cumulative medication when not used for consecutive days.
  - Expression of PRN ("as needed") medication orders—Measure logic is limited in representing the variability in PRN orders
  - The Lantana team recommends that the three other categories of the feasibility assessment take place as part of the beta testing phase of this measure and suggests AAN review the noted value set concepts to improve their specificity. New code requests or post-coordinated concepts may be an option to express detailed clinical information. This further exploration of feasibility will support the successful adoption and implementation of the measure."

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery? Is there any additional advise you can provide to the development team to improve the feasibility and thus the eventual reliability of this measure?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems*

and sites?

Preliminary rating for feasibility: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Low

**While the data in the EHR (assuming all entered appropriately) could be easily extracted, there are lots of questions about the criteria - and thus the accuracy of those data. Nonetheless, this is a major problem and it may be appropriate to start somewhere, even if that somewhat appears to have many flaws.

**weak definition of "opioid overuse" and insufficient granularity of data collection re. relationship between individual patients and treatment decisions.

** As indicated by the Lantana assessment, the required data elements are inconsistently documented as part of routine clinical care and, likely, rarely coded in the diagnoses. It would appear there would often be missing data elements, if this measure were to be implemented. Agree with the preliminary rating of low.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure [from OPUS]

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- The only current use of the measure reported by the developers is the American Academy of Neurology Axon Registry. Currently an internal benchmarking quality registry with plans to expand to external benchmarking. This tool enables neurology practices to identify and improve gaps in the quality of neurologic care. Axon Registry was launched in Q3 of 2015. The Overuse of Opioid Containing Medications for Primary Headache Disorders measure will be incorporated into the registry in 2016. Currently there are 3 pilot cohorts in progress providing data to the registry. Axon Registry is currently being piloted and will be available to all AAN US members. Currently, there are 42 practices (487 neurology providers) in the 3 pilot cohorts.
- This measure has been submitted to CMS for consideration for use in Physician Quality Reporting System (PQRS) and Merit-based Incentive Payment Systems (MIPS). Additionally, the AAN continues to outreach and partner with private payers to implement AAN developed measures in their payment programs.

Improvement results

- Developer states that data collection through the AAN´s Axon Registry was initiated in Q3 of 2015. Data validation is ongoing, and further analysis of performance is not available at this time.

Potential harms: The developer did not identify any unintended consequences related to this measure.

Feedback :

- This measure was submitted for MAP review for the Medicare Shared Savings Program in 2015. MAP determined that as a clinician-only level measure, this measure did not meet objectives for MSSP. Also in 2015,

this measure was reviewed for Physician Quality Reporting System, Physician Compare, Physician Feedback, and Value-Based Payment Modifier programs.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **In pilot program with neurology practices

**This could be very usable - even with all the issues noted above. As mentioned, I'd prefer the positive vs. the negative approach (frequency of use of the better/appropriate treatments). This could bypass some of the issues of arbitrary criteria.

**At the present time, the benefits of this measure do 'not' outweigh any potential unintended consequences.

** Care would need to be exercised in determination of how or if these results should be reported publically. It could be imagined that they could be used by opioid seekers to identify prescribers who were more likely to use opioids in the treatment of headache.

Criterion 5: Related and Competing Measures

Related or competing measures

None

Harmonization

N/A

Pre-meeting public and member comments

Comment by Amy Elaine Sanders, MD

Organization American Academy of Neurology

Comment #5572: This is an extremely important measure, with an enormous gap in care. EDs especially but also many physicians routinely prescribe opioids. Yet opioids are largely ineffective in controlling pain and are often the precipitant of rebound headache, which takes an already bad problem and makes it worse.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Overuse of Opioid Containing Medications for Primary Headache Disorders
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 1/15/2016
Additional information submitted 3/29/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome

☐ Health outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the PRO
PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Assessing patients for opioid containing medication overuse and treated or referred for treatment

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Triptans and ergots are considered first line acute treatments for migraine, not opioids or barbiturates by the US Headache Consortium Guideline.(1) The use of barbiturates or opioids increases the risk of chronic daily headache and drug induced hyperalgesia.(2) In one study, any use of barbiturates and opiates was associated with increased risk of transformed migraine after adjusting for covariates, while triptans were not.(3) In a sample of 5,796 people with headache, 4,076 (70.3%) were opioid nonusers, 798 (13.8%) were previous users, and 922 (15.9%) were current opioid users. Rates of headache-related health-care resource utilization were higher for all opioid-use groups for emergency department/urgent care, primary care, and specialty care visits compared to nonusers.(4) Opioid use for migraine is associated with more severe headache related disability, symptomology, comorbidities (depression, anxiety, and cardiovascular disease and events), and greater health-care resource utilization for headache. (4)

- Average headache frequency in days per month was higher among current opioid users. (4)
- Previous and current opioid users "had significantly higher average Migraine Disability Assessment (MIDAS) scores when compared to nonusers" . (4)
- Comorbidities tended to occur at higher rates among all opioid-use groups compared to nonusers, with depression significantly more common among opioid-use groups compared to nonusers . (4)
- Health-care resource use was higher among previous and current opioid user. (4)

Using the recommended first-line treatments for migraine would provide superior pain relief for migraine sufferers and reduce overuse of chronic daily headaches.

The 10 days/month for 3 months is diagnostic criteria taken directly from the International Classification of Headache Disorders, 3rd Edition. To meet the diagnostic criteria for opioid-overuse headache, a screening would be needed to confirm they meet these diagnostic criteria.

"8.2.4 Opioid-overuse headache
Diagnostic criteria:
1. Headache fulfilling criteria for 8.2 Medication-overuse Headache

2. Regular intake of one or more opioids1 on 10 days per month for >3 months."

"8.2 Medication-overuse headache (MOH)
Description:
Headache occurring on 15 or more days per month developing as a consequence of regular overuse of acute or symptomatic headache medication (on 10 or more, or 15 or more days per month, depending on the medication) for more than 3 months. It usually, but notinvariably, resolves after the overuse is stopped.
General comment:
In the criteria set out below for the various subtypes, the specified numbers of days of medication use considered to constitute overuse are based on expert opinion rather than on formal evidence."

Additional supporting literature addressing medication overuse headache and treatment are attached and include:

- Lipton RB. Risk Factors for and Management of Medication-Overuse Headache. 2015
- Giaberardino MA, Mitsikostas DD, Martelletti P. Update on Medication-Overuse Headache and Its Treatment. 2015
- Kristofferson ES and Lundqvist C. Medication-overuse headache: epidemiology, diagnosis and treatment. 2014
- Evers S and Jensen R. Treatment of medication overuse headache – guideline of the EFNS headache panel. 2011
- Grande RB, Aaseth K, Benth JS, et al. Reduction in medication-overuse headache after short information. The Akershus study of chronic headache. 2011
- Fritsche G, Frettloh J, Huppe M, et al. Prevention of medication overuse in patients with migraine. 2010
- Bigal ME, Rapopon AM, Sheftell FD, et al. Transformed migraine and medication overuse in a tertiary headache centre- clinical characteristics and treatment outcomes. 2004
- Limmroth V, Katsarava Z, Fritsche G, et al. Features of medication overuse headache following overuse of different acute headache drugs. 2002
- Hering R and Steiner TJ. Abrupt outpatient withdrawal of medication in analgesic-abusing migraineurs. 1991
- Granella F, Farina S, Malferrari G, et al. Drug abuse in chronic headache: a clinic-epidemiologic study.

The AAN is unaware of any guidelines developed that address referral and treatment to specialty care as such guideline statements could be viewed as self-serving with inherent bias from their creators. The need for specialty treatment following the identification of an opioid abuse concern is standard care. It is often in the patient's best interest to be referred to a provider who specializes in persistent headache, a provider who specializes in drug treatment and behavioral therapies, or detoxification center, such a referral must be personalized to the individual patient's situation. Although not specific to headaches, the CDC released their final guidelines on March 15, 2016 for chronic pain which includes the Class A recommendation "Clinicians should offer or arrange evidence-based treatment (usually medication-assisted treatment with buprenorphine or methadone in combination with behavioral therapies) for patients with opioid use disorder." It is noted most providers are not equipped to provide buprenorphine or methadone with behavioral therapies warranting a referral to treatment. These positions are held by the American Headache Society, the International Headache Society, and the American Academy of Neurology, and were signed off by all the member organization representatives that were part of the AAN headache quality measures development project.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

_____

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

A. National Institute for Health and Clinical Excellence (NICE). Headaches: Diagnosis and management of headaches in young people and adults. National Clinical Guideline Centre on behalf of the National Institute for Health and Clinical Excellence (NICE). September 2012; NICE clinical guideline 150. https://www.nice.org.uk/guidance/cg150/evidence/full-guideline-188258221

B. US Headache Consortium Matchar D, Young W, Rosenberg J et al. Evidence-Based Guidelines for Migraine Headache in the Primary Care Setting: Pharmacological Management of Acute Attacks Neurology 2000 http://tools.aan.com/professionals/practice/pdfs/gl0087.pdf

C. Lander-Gould A, Anderson W, Armstrong M et al. The American Academy of Neurology's Top Five Choosing Wisely recommendations. Neurology 2013; Published online before print February 20, 2013, doi: 10.1212/WNL.0b013e31828aab14 Neurology 10.1212/WNL.0b013e31828aab14 http://www.neurology.org/content/early/2013/02/20/WNL.0b013e31828aab14.full.pdf+html

D. Saper JR, Lake AE 34d. Continuous opioid therapy (COT) is rarely advisable for refractory chronic daily headache: Limited efficacy, risks, and proposed guidelines. Headache 2008; 48: 838-849 http://onlinelibrary.wiley.com/doi/10.1111/j.1526-4610.2008.01153.x/pdf

E. Department of Veteran Affairs and Department of Defense. VA/DoD Clinical Practice Guideline for Management of Opioids for Chronic Pain. 2010. Available at: http://www.healthquality.va.gov/Chronic_Opioid_Therapy_COT.asp

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

A. 1.3.8 Do not offer opioids for the acute treatment of tension-type headache

B. Oral opiate combinations may be considered for use in acute migraine when sedation side effects will not put the patient at risk and/or the risk for abuse has been addressed.

Parenteral opiates may be considered for rescue therapy in a supervised setting for acute migraine when sedation side effects will not put the patient at risk and when the risk abuse has been addressed.

C. Don't use opioid or butalbital treatment for migraine except as a last resort

D. Headache and other problems: Opioids are not usually indicated for migraine or TTH, or for patients with functional gastrointestinal problems. (No level of evidence noted)

E. Refer patients with significant headache to a neurologist for evaluation and treatment. (No level of evidence noted)

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

A. No grade

B. Grade A
Grade B

C. No grade

D. No grade

E. Grade B

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

A. N/A

B.

A. Multiple well-designed randomized clinical trials, directly relevant to the recommendation, yielded a consistent pattern of findings.

B. Some evidence from randomized clinical trials supported the recommendation, but the scientific support was not optimal. For instance, either few randomized trials existed, the trials that did exist were somewhat inconsistent, or the trials were not directly relevant to the recommendation. An example of the last point would be the case where trials were conducted using a study group that differed from the target group for the recommendation.

C. The US Headache Consortium achieved consensus on the recommendation in the absence of relevant randomized controlled trials.

C. N/A

D. TBD

E. The development process of this guideline follows a systematic approach described in "Guideline-forGuidelines," an internal working document of the VA/DoD Evidence-Based Practice Working Group that requires an ongoing review of the work in progress. Appendix A clearly describes the guideline development process followed for this guideline. In completing this OT guideline update, the WG relied heavily on the following evidence-based guideline: Chou R, Fanciullo GJ, Fine PG, et al. Clinical Guidelines for the Use of Chronic Opioid Therapy in Chronic Noncancer Pain. (APS/AAPM) The Journal of Pain 2009(Feb); 10(2):113-230. The WG reviewed the APS/AAPM 2009 guideline and made the decision to adopt several of their recommendations. The Working Group developed a revised comprehensive clinical algorithm that incorporates the assessment and determination of the appropriateness of OT as well as the management of therapy. Additional recommendations were added addressing treatment of specific adverse effects and for the diagnosis and management of aberrant behaviors that the Working Group considered to be of importance to patients in the healthcare systems of the VA and DoD.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

A. https://www.nice.org.uk/guidance/cg150/evidence/addendum-188258223 (Page 10)
B. http://tools.aan.com/professionals/practice/pdfs/gl0087.pdf (Page 55)
C. http://www.neurology.org/content/early/2013/02/20/WNL.0b013e31828aab14.abstract?sid=1d5b047e-12f0-4592-b460-098f78f2d078
D. http://www.readcube.com/articles/10.1111%2Fj.1526-4610.2008.01153.x
E. http://www.va.gov/painmanagement/docs/cpg_opioidtherapy_summary.pdf (Page 6)

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***
☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION
1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

_____
1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE
1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

_____
1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE
If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

A. In people with chronic or episodic migraine (with or without aura), what is the clinical evidence and cost-effectiveness of prophylactic pharmacological treatment with:
   - ACE (angiotensin converting enzyme) inhibitors and angiotensin II receptor antagonists
   - Antidepressants (SNRIs, SSRIs, tricyclics)
   - Centrally acting alpha-adrenergic-receptor agonists
   - Beta blockers
   - Calcium channel blockers
   - Antiepileptics
   - Other serotonergic modulators
   - NMDA receptor antagonists

   Outcomes considered were: change in migraine/headache days, 50% responder, and change in migraine/headache severity, change in migraine frequency, quality of life, change in acute medication use and serious adverse events.
B. The objective of the US Headache Consortium is to develop scientifically sound, clinically relevant practice guidelines on chronic headache for the primary care setting. This specific Guideline reviews the pharmacological treatment of acute migraine attacks.
   The US Headache Consortium identified the following goals of long-term migraine treatment:

· reduce attack frequency and severity,
· reduce disability,
· improve quality of life,
· prevent headache,
· avoid headache medication escalation, and
· educate and enable patients to manage their disease.

C. N/A

D. Summarizes the experience with COT for intractable headache, including the transformation from optimism about the potential to the current position that COT should rarely be administered to headache patients. COT should only be considered for those who meet strict criteria, and following rigorous procedures to assess response, adherence, and adverse effects. Moreover, prior to initiating COT, the prescribing physician should detail an explicit plan with the patient for opioid discontinuation if and when this becomes appropriate or necessary.

E. Scope of the Guideline:
• Offers best practice advice on the care of adults who may benefit from OT
• Addresses assessment and evaluation of chronic pain and appropriateness of OT
• Discusses primary intervention, referral, consultation and shared care in OT
• Addresses initiation, titration and maintenance of OT
• Presents and discusses formal treatment plans and treatment agreements for OT
• Presents updated pharmacotherapy advice on opioid medications that are FDA approved
• Provides guidance on assessing response to treatment, and determinations of adherence or abuse (aberrant drug-related behaviors)
• Addresses discontinuation of opioid therapy and follow-up • Discusses potential outcomes
• Does not address the use of opioids for patients receiving end of life treatment

1a.7.2. Grade assigned for the quality of the quoted evidence <u>with definition</u> of the grade:

A. There was no evidence identified for the effectiveness of this evidence, the recommendation is based on the absence of evidence and GDG informal consensus.

1. A quality rating was assigned, based on the study design. RCTs start HIGH and observational studies as LOW, uncontrolled case series as LOW or VERY LOW.
2. The rating was then downgraded for the specified criteria: Study limitations, inconsistency, indirectness, imprecision and reporting bias. These criteria are detailed below. Observational studies were upgraded if there was: a large magnitude of effect, dose-response gradient, and if all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results showed no effect. Each quality element considered to have 'serious' or 'very serious' risk of bias were rated down -1 or -2 points respectively.
3. The downgraded/upgraded marks were then summed and the overall quality rating was revised. For example, all RCTs started as HIGH and the overall quality became MODERATE, LOW or VERY LOW if 1, 2 or 3 points were deducted respectively.

B. Grade A

A. Multiple well-designed randomized clinical trials, directly relevant to the recommendation, yielded a consistent pattern of findings.

B. Some evidence from randomized clinical trials supported the recommendation, but the scientific support was not optimal. For instance, either few randomized trials existed, the trials that did exist were somewhat inconsistent, or the trials were not directly relevant to the recommendation. An example of the last point would be the case where trials were conducted using a study group that differed from the target group for the recommendation.

C. The US Headache Consortium achieved consensus on the recommendation in the absence of relevant randomized controlled trials.

D. No grades given

E.

A. A strong recommendation that clinicians provide the intervention to eligible patients. Good evidence was found that the intervention improves important health outcomes and concludes that benefits substantially outweigh harm.
B. A recommendation that clinicians provide (the service) to eligible patients. At least fair evidence was found that the intervention improves health outcomes and concludes that benefits outweigh harm.
C. No recommendation for or against the routine provision of the intervention is made. At least fair evidence was found that the intervention can improve health outcomes, but concludes that the balance of benefits and harms is too close to justify a general recommendation.
D. Recommendation is made against routinely providing the intervention to asymptomatic patients. At least fair evidence was found that the intervention is ineffective or that harms outweigh benefits.
I. The conclusion is that the evidence is insufficient to recommend for or against routinely providing the intervention. Evidence that the intervention is effective is lacking, or poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: A. Dates not given. B. Dates not given. D. Through May 2005. E. January 2003 to March 2009.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

A. The recommendation is based on the absence of evidence and GDG informal consensus
B. TBD
C. TBD
D. Included were 41 randomized trials.
E. 4 studies, Class 1, fair quality of evidence, Grade B recommendation.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

A. The recommendation is based on the absence of evidence and GDG informal consensus
B. Recommendations ranged from Levels A to C depending on the specific opiate recommendation.
C. TBD
D. Article does not describe effects, imprecision, etc. Most studies had high drop-out rates suggesting these studies are not high quality.
E. 4 studies, Class 1, fair quality of evidence, Grade B recommendation.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

A. The GDG agreed that pain free at 2 hours was the most important outcome. There is no evidence for the effectiveness of opioids in the acute treatment of tension type headache. GDG informal consensus agreed that there are considerable recognized side effects of opioids including an increased risk of medication overuse headache and therefore their use should not be recommended.
B. Document did not discussion estimates of benefit or magnitude and direction of effect on outcome.
C. TBD
D. One meta analysis found that mean pain relief with opioids was about 30%, with only a minority appearing to benefit from long-term treatment. In one study reviewed, only 20% of patients had still had relief with oral morphine after one year. Five studies found no significant difference between opioid or placebo treatment. Despite some benefit from short-term use of opioids, only 44% of 388 patients who were offered continued COT elected to remain on opioids for periods ranging between 7 and 24 months. Others have subsequently noted that "The evidence base for this type of pain management is meager because the needed randomized controlled trials, which ideally should last for several years, have not been performed".Another meta analysis concluded that weak evidence suggests that oral and intrathecal opioids reduce pain long-term for patients who benefit short-term with minimal adverse effects. There was insufficient data from transdermal studies to quantify pain relief.
E. Document did not discussion estimates of benefit or magnitude and direction of effect on outcome.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

A. No studies reported outcome data for time to freedom from pain, headache response at up to 2 hours, headache response at 24 hours, freedom from pain at 24 hours, functional health outcomes, or incidence of serious adverse events. Study did not do a specific benefits vs. harms analysis.
B. The oral opiate analgesics reviewed were associated with a higher rate of adverse events than was placebo, but were similar to aspirin and better than ergotamine in that respect. The most commonly reported adverse events included dizziness, fatigue, nausea, and drowsiness. Adverse events were much more frequently reported with nasal butorphanol than with placebo or with oral opiate analgesics. Study did not do a specific benefits vs. harms analysis.
C. N/A
D. The most common adverse effect stemming from opioid administration is nausea, sometimes associated with emesis, occurring in 26% of 8855 patients in a retrospective cohort study of patients receiving short-term opioids in 35 community-based and tertiary hospitals. In one meta-analysis constipation and nausea were statistically significant.
E. Opioid therapy is a therapeutic trial. Prior to such a trial, the provider should determine that the potential benefits are likely to outweigh the potential harms, and the patient should be fully informed and should consent to the therapy. Opioid therapy should be tapered off and discontinued if real or potential harms outweigh real or potential benefits.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

A. No new studies on the use of opioids were included in the updated review of evidence NICE conducted in 2014. https://www.nice.org.uk/guidance/cg150/evidence/evidence-update-october-2014-2186959789
B. No new studies on the use of opioids that would change the recommendatons – document still stands as is.
C. N/A
D. No new studies on the use of opioids that would change this article.
E. No new studies on the use of opioids that would change the recommendations – document still stands as is.

1. On 3/22/216 the FDA announced required class-wide safety labeling changes for immediate-release (IR) opioid pain medications. Among changes, the FDA is requiring a new black boxed warning about the serious risks of misuse, abuse, addiction, overdose and death. The FDA is also requiring several additional safety labeling changes across all prescription opioid products to include additional information on the risk of these medications. These changes are part of the FDA's effort to inform prescribers about the importance of balancing the serious risks of opioids with their role in managing pain. Direct quote from Robert Califf, MD, FDA commissioner: "Opioid addiction and overdose have reached epidemic levels over the past decade, and the FDA remains steadfast in our commitment to do our part to help reverse the devastating impact of the misuse and abuse of prescription opioids." FDA Press Release. March 22, 2016. Available at: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm491739.htm
2. On 3/15/16, the CDC released their CDC Guidelines for Prescribing Opioids for Chronic Pain 2016, which includes the following Class A recommendation:
    a. **Nonpharmacologic therapy and nonopioid pharmacologic therapy are preferred for chronic pain. Clinicians should consider opioid therapy only if expected benefits for both pain and function are anticipated to outweigh risks to the patient. If opioids are used, they should be combined with nonpharmacologic therapy and nonopioid pharmacologic therapy, as appropriate (recommendation category: A, evidence type: 3).**
    b. **Before starting and periodically during opioid therapy, clinicians should discuss with patients known risks and realistic benefits of opioid therapy and patient and clinician responsibilities for managing therapy (recommendation category: A, evidence type: 3).**
    c. **Before starting and periodically during continuation of opioid therapy, clinicians should evaluate risk factors for opioid-related harms. Clinicians should incorporate into the management plan strategies to mitigate risk, including considering offering naloxone when factors that increase risk for opioid overdose, such as history of overdose, history of substance use disorder, higher opioid dosages (≥50 MME/day), or concurrent benzodiazepine use, are present (recommendation category: A, evidence type: 4).**
    d. **Clinicians should avoid prescribing opioid pain medication and benzodiazepines concurrently whenever possible (recommendation category: A, evidence type: 3).**
    e. **Clinicians should offer or arrange evidence-based treatment (usually medication-assisted treatment with buprenorphine or methadone in combination with behavioral therapies) for patients with opioid use disorder (recommendation category: A, evidence type: 2).**

    The report includes the following as well: "… In some clinical contexts (e.g., headache or fibromyalgia), expected benefits of initiating opioids are unlikely to outweigh risks regardless of previous nonpharmacologic and nonopioid pharmacologic therapies used." Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain — United States, 2016. MMWR Recomm Rep 2016;65:1–49. DOI: http://dx.doi.org/10.15585/mmwr.rr6501e1.
3. The American Headache Society released 5 Choosing Wisely statements in November/December 2013 which includes the statement: "(4) don't prescribe opioid- or butalbital-containing medications as a first-line treatment for recurrent headache disorders…" Choosing Wisely statements are intended as a starting point of treatment conversations for patients and providers. This statement supports that opioid prescription use for primary headache is not the standard of care. Loder E, Weizenbaum E, Frishberg B, Silberstein S; American Headache Society Choosing Wisely Task Force. Choosing wisely in headache medicine: the American Headache Society's list of five things physicians and patients should question. Headache 2013;53:1651–9

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

F. Choosing Wisely mirrors the guideline development process. A narrow literature review was conducted.

1a.8.2. Provide the citation and summary for each piece of evidence.

C.

Silberstein SD; US Headache Consortium. Practice parameter: Evidence-based guidelines for migraine headache (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology [Internet]. 2000;55(6):754-762.

See information under B for summary of this article.

Evers S, Afra J, Frese A, Goadsby PJ, Linde M, May A, Sandor PS, European Federation of Neurological Societies. EFNS guideline on the drug treatment of migraine – revised report of an EFNS task force. Eur J Neurol [Internet]. 2009 Sep;16(9):968-81.

Summary: Opioids are of only minor efficacy, no modern controlled trials are available for these substances.

Institute for Clinical Systems Improvement. Headache, Diagnosis and Treatment of (Guideline) [Internet]. Bloomington, MN: Institute for Clinical Systems Improvement; 2011 [cited 2012 Oct 25]. Available from: www.icsi.org/headache/headache__diagnosis_and_treatment_of_2609.html.

Summary: Clinicians may manage mild migraines with over-the-counter medications. Clinicians may use triptans for mild migraine pain levels. Clinicians should avoid the use of opiates and barbiturates in the treatment of headache. In general, opiates are characterized by having a short pain-relief window, release inflammatory neurochemicals, and increase vasodilation; none of these addresses the currently known treatment issues and pathophysiology of migraine.

1. FDA Press Release. March 22, 2016. Available at: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm491739.htm
2. ." Dowell D, Haegerich TM, Chou R. CDC Guideline for Prescribing Opioids for Chronic Pain — United States, 2016. MMWR Recomm Rep 2016;65:1–49. DOI: http://dx.doi.org/10.15585/mmwr.rr6501e1.
3. Loder E, Weizenbaum E, Frishberg B, Silberstein S; American Headache Society Choosing Wisely Task Force. Choosing wisely in headache medicine: the American Headache Society's list of five things physicians and patients should question. Headache 2013;53:1651–9

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
HA_MeasSubm_Evidence_2016-01-15.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Triptans and ergots are considered first line acute treatments for migraine, not opioids or barbiturates by the US Headache Consortium Guideline. The use of barbituates or opioids increases the risk of chronic daily headache and drug induced hyperalgesia. In one study, any use of barbituates and opiates was associated with increased risk of transformed migraine after adjusting for covariates, while triptans were not. In a sample of 5, 796 people with headache, 4,076 (70.3%) were opioid nonusers, 798 (13.8%) were previous users, and 922(15.9%) were current opioid users.

Rates of headache-related health-care resource utilization were higher for all opioid-use groups for emergency department/urgent care, primary care, and specialty care visits compared to nonusers. Opioid use for migraine is associated with more severe headache-related disability, symptomology, comorbidities (depression, axiety, and cardiovascular disease and events), and greater health-care resource utilization for headache.
- Average headache frequency in days per month was higher among current opioid users
- Previous and current opioid users "had significantly higher average Migraine Disability Assessment (MIDAS) scores when compared to nonusers.
- Comorbidities tended to occur at higher rates among all opioid-use groups compared to nonusers, with depression significantly more common among opioid-use groups compared to nonusers.
- Health-care resource use was higher among previous and current opioid user

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
Using the recommended first-line treatments for migraine would provide superior pain relief for migraine sufferers and reduce overuse of chronic daily headaches.

1. Matchar DB, Young WB, Rosenberg J, et al. Multispecialty consensus on diagnosis and treatment of headache: pharmacological management of acute attacks. Available at http://www.aan.com/professionals/practice/pdfs/g10087.pdf
2. Lipton RB, Buse DC, Serrano D, et al. Examination of Unmet Treatment Needs Among Persons with Episodic Migraine: Results of the American Migraine Prevalence and Prevention Study
3. Bigal ME, Serrano D, Buse D, et al. Acute migraine medications and evoluation from episodic to chronic migraine: a longitudinal population-based study. Headache 2008; 48(8):1157-68
4. Buse DC, Pearlman SH, Reed ML, et al. Opioid Use and Dependence among Persons with Migraine: Results of the AMPP Study Headache. The Jouranl of Head and Face Pain 2012; 52(1):18-36.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity,**

gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Data analysis on disparities from the measure as specified has not been conducted at this time.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Migraine ranks in the top 20 of the world´s most disabling medical illnesses. In US population studies, the prevalence of migraine is approximately 18% in women and 6% in men. The prevalence of migraine in children is 7.7%. Migraines affect 29.5 million Americans.

Headaches, notably migraine, are associated with significant disability. Migraine is strongly associated with anxiety and mood disorders, chronic pain disorders, and epilepsy. Headache can be attributed to comorbid conditions which need adequate management. Migraine is associated with increased risk for another physical and psychiatric comorbidities and this risk increases with headache frequency.

90% of people with headache have some headache-related disability, and approximately half are severely disabled or require bed rest. 9 out of 10 people with headache report they can´t "function normally" during days in which a migraine strokes and 3 in 10 require bed rest. More than 25% of migraine sufferers missed at least one day of work over the past three months due to a migraine.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Migraine ranks in the top 20 of the world´s most disabling medical illnesses. In US population studies, the prevalence of migraine is approximately 18% in women and 6% in men. The prevalence of migraine in children is 7.7%. Migraines affect 29.5 million Americans.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
World Health Organization. Headache Disorders Fact Sheet.
http://www.who.int/mediacentre/factsheets/fs277/en/ Accessed. 8.22.2013

Alliance for Headache Disorders Advocacy. Headache Disorders Fact Sheet
www.allianceforheadacheadvocacy.org Accessed 8.24.2013

Leonardi M, Musicco M, Nappi G. Headache as a major public health problem: Current status. Cephalalgia. 1998: 18 (S21)::66-69)

National Headache Foundation Impact of Migraine: Evaluation Patient Disability
http://www.headaches.org/pdf/Monograph12.pdf

Smitherman TA, Burch R, Sheikh H, Loder E. The prevalence, impact, and treatment of migraine and severe headache in the United States: a review of statistics from national surveillance studies. Headache. 2013. 53(3):427-36.

International Association for the Study of Pain. Epidemiology of Headache Fact Sheet 2012. /www.iasppain.org/

Smitherman TA, Burch R, Sheikh H, Loder E. The prevalence, impact, and treatment of migraine and severe

headaches in the United States: a review of statistics from national surveillance studies. Headache. 2013 Mar; 53(3):427-36. doi: 10.1111/head.12074. Epub 2013 Mar 7

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not Applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Behavioral Health : Alcohol, Substance Use/Abuse, Mental Health : Alcohol, Substance Use/Abuse, Neurology

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Overuse, Safety : Medication Safety

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.aan.com/practice/quality-measures/

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is an eMeasure **Attachment:** OpioidOverusePrimaryHeadache_v4_Artifacts_07202015_Bonnie.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** OpioidOverusePrimaryHeadache_v4_Fri_Jul_24_10.56.36_CDT_2015.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not applicable.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients assessed for opioid containing medication overuse within the 12-month measurement period and treated or referred for treatment if identified as overusing opioid containing medication which is defined as:
Using opioid containing medication for greater than or equal to 10 days per month for more than 3 months.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
12 months

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Measure is not applicable to claims reporting, and values for gathering information through EHR and Registry sources is attached in S.2b.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients aged 12 years and older diagnosed with a primary headache disorder* and taking opioid containing medication.

*Define Primary Headache: A headache that is not caused by another disease or medical condition. For the purpose of this measure

this includes the following types of headache:
Migraine - Migraine without aura, migraine with aura, childhood periodic syndromes that are commonly precursors of migraine, retinal migraine, complications of migraine, probable migraine.
Tension-Type Headache (TTH) - Infrequent episodic TTH, frequent episodic TTH headache, chronic TTH, probable TTH
Cluster Headache (CH) and Other Trigeminal Autonomic Cephalgias: Cluster headache, paraxysmal hemicrania, short-lasting unilateral neuralgia from headache attacks with conjuctival injection and tearing (SUNCT), probably trigeminal autonomic cephalgia
Other Primary Headaches: Primary stabbing headache, primary cough headache, primary exertional headache, primary headache associated with sexual activity, hypnic headache, primary thunderclap headache, hemicrania continua, new daily-persistent headache.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Children's Health, Populations at Risk

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Measure is not applicable to claims reporting, and values for gathering information through EHR and Registry sources is attached in S.2b.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
No Exclusions.

Medical exceptions for not assessing, treating, or referring patient for treatment of opioid medication overuse include:
Patient already assessed and treated for opioid use disorder within the last year.
Patient has a documented failure of non-opioid options and is not identified as overusing opioid containing medication.
Patient has contraindications to all other medications for primary headache.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Measure is not applicable to claims reporting, and values for gathering information through EHR and Registry sources is attached in S.2b.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. **Type of score:**

Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
See S.2a for details

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
For use in the American Academy of Neurology (AAN) Axon Registry only - If data is missing from denominator, the case is deleted. If data met for denominator then case information is included for the measure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The American Academy of Neurology (AAN) Axon Registry planned for 2016

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Individual

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Ambulatory Care : Urgent Care, Emergency Medical Services/Ambulance, Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not Applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form

testing_form_for_trial_use_AAN_Opioids.docx

National Quality Forum

Measure Testing Form for Trial Approval Program

Measure Title: Overuse of Opioid Containing Medications for Primary Headache

Date of Submission: January 15, 2016

Type of Measure:

☐ Composite

☐ Outcome (*including PRO-PM*)

☐ Cost/resource

☐ Process

☐ Efficiency

☐ Structure

Instructions

- A measure submission that is to be considered for the Trial Approval Program must complete this form in its entirety. Either a test data set provided by the measure developer, or the use of the Bonnie tool is acceptable to provide preliminary testing results,
- **For <u>all</u> measures being submitted for potential acceptance into the Trial Approval Program, each section <u>must be filled out as completely as possible.</u>**
- Respond to <u>all</u> questions as instructed with answers immediately following the question. All information on testing of either a sample data set or results from Bonnie testing that can demonstrate, to the extent possible, the the measure meets the reliability and validity must be in this form..
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions at trialmeasures@qualityforum.org

<u>DATA and SAMPLING INFORMATION</u>

1. DATA/SAMPLE USED FOR <u>PRELMINARY</u> TESTING OF THIS MEASURE

It is important that the measure developer use a data set to conduct preliminary testing in order to evaluate the measure logic and the inclusions/exclusions for the population used in the measure.

- **What type of data was used for testing**? (*The measure developer must provide a test data set that will provide some initial information to be used for the evaluation, or the Bonnie testing tool can use can be used to create a sample data set using synthesized patients.)* Please indicate whether the test data set used was provided through the measure developer, or through the Bonnie tool.

Bonnie

- **If Bonnie was <u>NOT</u> used, please identify the specifications for the test dataset** (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured)*

- **What levels of analysis were tested (either through the test data set or Bonnie)**? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*) *in order to determine its suitability for inclusion into the Trial Approval Program.,*

Individual clinician

Measure Specified to Measure Performance of:

Measure Tested at Level of:

☐ individual clinician

☐individual clinician

☐ group/practice

☐ group/practice

☐ hospital/facility/agency

☐ hospital/facility/agency

☐ other: Click here to describe

☐ other: Click here to describe

1.4. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis)*

36 test cases in the Bonnie

1.5. Please refer to the guidance for Bonnie testing found at this link. Bonnie testing results may be compiled into spreadsheet or table, which must be completed in its entirety, to the extent possible, in order to provide a basis for evaluation to determine the acceptability of the measure for inclusion in the Trial Approval program. *Any questions regarding the completion of this form can be directed to NQF Staff at trialmeasures@qualityforum.org.*

<u>RELIABILITY AND VALIDITY ASSESSMENTS</u>

Note: The information provided in this next section is intended to aid the Standing Committee and other stakeholders in understanding to what degree the measure is both reliable and valid. While it is not possible to provide comprehensive results due to the lack of actual testing data, the developer needs to provide as much information as possible based on their interpretation of the results from the sample test data.

2.1 Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score. **What is your interpretation of the results in terms of demonstrating reliability**? (i.*e., what do the sample results mean and what are the norms for the test conducted?*) Please summarize the plan for future testing of reliability if the measure is accepted into the Trial Approval Program. Include descriptions of:

- Inter-abstractor reliability, and data element reliability of all critical data elements
- Computation of the performance measure score (e.g., signal-to-noise analysis)?

The AAN is currently working with Minnesota Community Measurement to test this measure for feasibility and reliability. The AAN will also utilize the Axon Registry to gather data collected directly from the participant's EHR to test for feasibility and reliability. The AAN's Axon Registry also constantly measures physician performance on this measure.

2.2 Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score. **What is your interpretation of the results in terms of demonstrating validity**? (i.*e., what do the results mean and what are the norms for the test conducted?*). Please summarize the plan for future testing of validity if the measure is accepted into the Trial Approval Program. Include the method(s) of validity testing and what it will test (describe the steps—do not just name a method; what will be tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis will be used used)

See above, validity will be tested by both Minnesota Community measurement and in the future through the AAN's Axon Registry.

2.3 Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion. **What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results?** (*i.e., the value outweighs the burden of increased data collection and analysis).* Please summarize the plan for future testing of exclusions if the measure is accepted into the Trial Approval Program. Describe the method of testing exclusions and what it will test (describe the steps—do not just name a method; what will be tested, e.g., whether exclusions affect overall performance scores; what statistical analysis will be used)

Exclusions and the burden of exclusions included in the measures are evaluated not only as the e-measures are developed but on an ongoing basis. The AAN has a system in place to garner user feedback from those members participating in the AAN's Axon Registry, burden of exclusions is an issue the AAN will continue to assess participants of the registry.

2.4 Risk Stratification (applicable ONLY to outcome or resource use measures). If an outcome or resource use measure will not be risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities. If risk adjustment/stratification is needed then please describe the conceptual/clinical and statistical methods and criteria that will be used to select patient factors (clinical factors or sociodemographic factors) that will be used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

N/A

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in a combination of electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment **Attachment:** Opioid_Overuse_Headache_Feasibility.docx

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

See attached Alpha Testing Results in Appendix. From report, "The Overuse of Opioid Containing Medications for Primary Headache Disorders measure passed all alpha testing, which relates to measure face validity, data element reliability, and measure logic testing. NQF recommends that an early feasibility assessment take place to determine data availability, data accuracy, data standards, and data workflow.

- Data availability—The extent to which the data are readily available in a structured format across EHR systems.
- Data accuracy—The extent to which the information contained in the data is correct. This includes whether the most accurate data source is used and/or captured by the most appropriate healthcare professional or recorder.
- Data standards—The extent to which the data element is coded using a nationally accepted terminology/vocabulary standard. Standard data elements, associated definitions and code sets, and mapping to the QDM are expected.
- Workflow—The extent to which the effort of capturing the data element interferes with a typical provider workflow.

A feasibility assessment conducted at the outset of the project addressed one of the four recommended categories for feasibility assessment: data standards.

Properly conducted eCQM testing and analysis is critical for endorsement by NQF and for success in implementing a measure. During reviews, Measure Analysts discussed issues related to feasibility with the Measure Steward. The Lantana team shared potential challenges in the measure related to date standards:

- Non-use codes for treatment—Currently this concept is not captured discretely in Behavioral Health Information Technology systems. According AAN, neurologists do not currently utilize any standard codes. Candidate for revision is Referral for Opioid Overuse Treatment SNOMEDCT Value Set; its associated OID is 2.16.840.1.113762.1.4.1111.90.
- Failed treatment—This procedure is not consistently documented or captured in EHRs. Candidate for revision is Treatment Failure SNOMEDCT Value Set; its associated OID is 2.16.840.1.113762.1.4.1111.93.
- Opioid assessments—There is no standard form or question that captures data related to whether an opioid assessment was conducted. Candidate for revision is Opioid Overuse Assessment SNOMEDCT Value Set; its associated OID is 2.16.840.1.113762.1.4.1111.89.
- Opioid overuse—The current definition of this concept captures only overuse fitting the precise description. Example: patient takes opioids three times a day, every day for 2 months, would not be considered as overusing.
- Calculating cumulative opioid use—The QDM is limited in its ability to calculate cumulative medication when not used for consecutive days.
- Expression of PRN ("as needed") medication orders—Measure logic is limited in representing the variability in PRN orders

The Lantana team recommends that the three other categories of the feasibility assessment take place as part of the beta testing phase of this measure and suggests AAN review the noted value set concepts to improve their specificity. New code requests or post-coordinated concepts may be an option to express detailed clinical information. This further exploration of feasibility will support the successful adoption and implementation of the measure."

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

There are no fees, licensing, or other requirements that impact use.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Payment Program Professional Certification or Recognition Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Quality Improvement (Internal to the specific organization) Axon Registry https://www.aan.com/practice/axon-registry/

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

The American Academy of Neurology Axon Registry. Currently an internal benchmarking quality registry with plans to expand to external benchmarking. This tool enables neurology practices to identify and improve gaps in the quality of neurologic care. Axon Registry was launched in Q3 of 2015. The Overuse of Opioid Containing Medications for Primary Headache Disorders measure will be incorporated into the registry in 2016. Currently there are 3 pilot cohorts in progress providing data to the registry. Axon Registry is currently being piloted and will be available to all AAN US members. Currently, there are 42 practices (487 neurology providers) in the 3 pilot cohorts.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

No policies or actions restrict the use of the measure in accountability or public reporting programs, and no known barriers to implementation. Measurement set released January 2015, and has not been implemented to date.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

This measure has been submitted to CMS for consideration for use in Physician Quality Reporting System (PQRS) and Merit-based Incentive Payment Systems (MIPS). Additionally, the AAN continues to outreach and partner with private payers to implement AAN developed measures in their payment programs.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Not available at this time. Data collection through the AAN´s Axon Registry was initiated in Q3 of 2015. Data validation is ongoing, and further analysis of performance is not available at this time.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not Applicable

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No unintended negative consequences have been identified at this time.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Not Applicable

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed**

measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not Applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Opioid_appendix.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Academy of Neurology
Co.2 **Point of Contact:** Erin, Lee, elee@aan.com, 612-928-6020-
Co.3 **Measure Developer if different from Measure Steward:** American Academy of Neurology
Co.4 **Point of Contact:** Erin, Lee, elee@aan.com, 612-928-6020-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Voting Work Group Members:
Stephen Ross, MD (American Academy of Neurology Project Co- Chair)
Eric Wall, MD (American Academy of Family Physicians Project Co- Chair)

American Academy of Neurology
Wayne Anderson, MD
David Roeltgen, MD
Stephen Silberstein, MD
American Academy of Family Physicians
Amy Schneider, MD
American Academy of Pediatrics
Thomas K. Koch, MD
American Academy of Physical Medicine and Rehabilitation
Thomas Watanabe, MD
American Board of Emergency Medicine
Robert P Wahl, MD
American College of Emergency Physicians
Andy Jagoda, MD
American College of Physicians
James Foody, MD
American College of Radiology
David Seidenwurm, MD
American Osteopathic Association
J. Mark Bailey, DO, PhD
American Physical Therapy Association
Shannon Petersen, PT
American Psychiatric Association
Jon Streltzer, MD
American Psychological Association
Robert Nicholson, PhD
Child Neurology Society
M. Cristina Victorio, MD

Alliance for Headache Disorders Advocacy
William B. Young, MD
American Headache Society
Elizabeth Loder, MD
National Headache Foundation
Mary Franklin

Non-Voting Work Group Members:
Charles Flippen II, MD (Work Group Facilitator)
Eric Cheng MD, MS (Work Group Facilitator)
Amy Sanders, MD, MS (Work Group Facilitator)
Rebecca J. Swain-Eng, MS, CAE (American Academy of Neurology Staff)
Gina Gjorvad (American Academy of Neurology Staff)
Becky Schierman, MPH (American Academy of Neurology Staff)
Elvia Chavarria, MPH (American Medical Association Physician Consortium for Performance Improvement Staff)

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2015

Ad.3 **Month and Year of most recent revision:** 01, 2015

Ad.4 **What is your frequency for review/update of this measure?** Every 3 years

Ad.5 **When is the next scheduled review/update for this measure?** 01, 2017

Ad.7 **Disclaimers:** Physician Quality Measures (Measures) and related data specifications developed by the American Academy of Neurology (AAN) are intended to facilitate quality improvement activities by physicians. These measures are intended to assist physicians in enhancing quality of care. Measures are designed for use by any physician who manages the care of a patient for a specific condition or for prevention. The AAN encourages use of these Measures by other health care professionals, as appropriate. These Measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. The AAN encourages testing and evaluation of its Measures.
Measures are subject to review and may be revised or rescinded at any time by the AAN. The measures may not be altered without prior written approval from the AAN. The measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes (e.g., use by health care providers in connection with their practices). Commercial use is defined as the sale, license, or distribution of the measures for commercial gain, or incorporation of the measures into a product or service that is sold, licensed, or distributed for commercial gain. Commercial uses of the measures require a license agreement between the user and the AAN. Neither the AAN nor its members shall be responsible for any use of the measures.

THESE MEASURES AND SPECIFICATIONS ARE PROVIDED “AS IS” WITHOUT WARRANTY OF
ANY KIND

Ad.8 **Additional Information/Comments:**

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2832
Measure Title: STK 02: Discharged on Antithrombotic Therapy
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of ischemic stroke patients prescribed antithrombotic therapy at hospital discharge.

This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-4: Thrombolytic Therapy,STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-6 Discharged on Statin Medication, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs. STK-2, Discharged on Antithrombotic Therapy, is one of six of the measures in this set that have been reengineered as eCQMs and are included in the EHR Incentive Program and Hospital Inpatient Quality Reporting Program.
Developer Rationale: Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. The effectiveness of antithrombotic agents in reducing stroke mortality, stroke-related morbidity and recurrence rates has been studied in several large clinical trials. Studies have shown that antithrombotic therapy prescribed at discharge following acute ischemic stroke reduces stroke mortality and morbidity.

Healthcare organizations that track antithrombotic therapy prescribed at discharge for internal quality improvement purposes have seen an improvement in the measure rate over time. The chart-abstracted measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to which will also promote improvements in quality at the national level.

Numerator Statement: Patients prescribed antithrombotic therapy at hospital discharge.
Denominator Statement: Patients with a principal diagnosis of ischemic stroke.
Denominator Exclusions: Patients with comfort measures documented.
Patients admitted for elective carotid intervention. This exclusion is implicitly modeled by only including non-elective hospitalizations.
Patients discharged to another hospital
Patients who left against medical advice
Patients who expired
Patients discharged to home for hospice care
Patients discharged to a health care facility for hospice care

Denominator Exceptions:
Patients with a documented reason for not prescribing antithrombotic therapy at discharge.

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Pharmacy
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary:

Evidence for this new eMeasure is the same as the existing measure #0435 STK 02: Discharged on Antithrombotic Therapy:

- The body of evidence consistently supports that prolonged antithrombotic therapy reduces the risk of death in patients with a prior history of ischemic stroke. Based on these findings, antithrombotic therapy should be prescribed at discharge following acute ischemic stroke, unless there are contraindications to therapy.
- 2010 AHA/ASA Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack, page 249.
    - Class I, Level of Evidence A - For patients with noncardioembolic ischemic stroke or TIA, the use of antiplatelet agents rather than oral anticoagulation is recommended to reduce the risk of recurrent stroke or another cardiovascular event.
    - Aspirin (50 mg/d to 325 mg/d) monotherapy (Class I, Level of Evidence A), the combination of aspirin 25 mg and extended-release dipyridamole 200 mg twice daily (Class I, Level of Evidence B), and clopidogrel 75 mg monotherapy daily (Class IIa, Level of Evidence B), are all acceptable options for initial therapy. The selection of an antiplatelet agent should be individualized on the basis of patient risk factor profiles, cost, tolerance, and other clinical characteristics.

Questions for the Committee:

- *The developer attests the underlying evidence for the measure #0435 (chart-abstracted version) has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the chart-abstracted measure has not changed and can be applied to the eMeasure, therefore there is no need for repeat discussion and voting on Evidence?*

Guidance from the Evidence Algorithm: Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency: high (Box 5) → Box 5a High

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**
Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer states no performance data for the electronic version of this measure are yet available.
- This eCQM is currently included in the Hospital Inpatient Quality Reporting Program (IQR) and EHR Incentive program.
    - In CY2016, CMS is requiring organizations participating in HIQR to electronically submit one quarter of data on 4 of 28 available eCQMs. This measure is 1 of the 28 eCQMs.

- The developer provided a summary of data from the literature that demonstrated underutilization of antithrombotic therapy and a performance gap in Medicare patients.
- The developer also provided performance data from the chart abstracted measure which seems to be very high with little to no room for improvement.

Disparities:

- The developer does not provide disparities data from use of this measure.
- Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age."

Questions for the Committee:

- *Is there a gap in care that warrants a national performance eMeasure?*
- *How can this measure be used to better understand disparities in care and outcomes for certain groups?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **No new studies/information

**This measure is the EHR measure -assesses same process as #0435-paper measure of antithrombotic therapy at discharge

**I am not aware of new data

1b. Performance Gap

Comments: **no performance data for e-version available; no disparities data

**Emeasure --no overall performance data vailable yet, In CY2016 CMS is requiring all organizations participating in HIQR to submit data for 1 quarter on 4 of 28 available emeasures (eCQMs). This is 1 of the 28 eCQMs.

**There is minimal performance gap

1c. High Priority (previously referred to as High Impact)

Comments: **NA

**Yes

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic Clinical Data, Electronic Health Record (EHR), Imaging/Diagnostic Study, Laboratory Data, and Pharmacy Data. This is an eMeasure.

Specifications:

- HQMF specifications for eMeasure are included in the document set on SharePoint. See eMeasure Technical Review below.
- The developer states that changes have been made to the eCQM specifications to reflect revisions to the chart

abstracted measure from which this measure is derived.

Questions for the Committee :

- o *Are all the data elements clearly defined? Are all appropriate codes included?*
- o *Is the logic or calculation algorithm clear?*
- o *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor review:

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously Demonstrated through Bonnie testing
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible. This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment.

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☒ **No**

Method(s) of reliability testing:

- Dataset used for testing included 22 synthetic records created in the Bonnie testing system. Measure not tested in EHR.
- Data element validity testing was performed and will count for data element reliability as well – see validity testing section.

Questions for the Committee:

- o *Is the test sample adequate to generalize for widespread implementation?*
- o *Do the results from the Bonnie tool demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm: Precise specifications (Box 1) →empiric reliability testing as specified (Box 2)

→ empiric validity testing of patient-level data (Box 3) → YES: Use rating from validity testing of patient-level data elements (Box 10)→Assessed the measure logic only (Box 11) →Low certainty or confidence data used in measure are valid (Box 12b)→Low (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Preliminary rating for reliability: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:

o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**
☒ **Empirical validity testing of the measure score**

Validity testing method:

- The Bonnie testing tool and environment with 22 synthetic patient records were used to test the measure logic and value sets. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure.
- The numerator, denominator, exclusions/exceptions , and logic statement were tested to confirm actual results met expectations.
- The developer also conducted a review of the measure specifications to confirm the measure logic was properly expressed with the current version of the QDM and that the logic matches the clinical intent of the measure.

Validity testing results:

- The testing results from the Bonnie tool reached 100% coverage and confirmed there was a test case for each pathway of logic (negative and positive test cases).
- The measure also had a 100% passing rate which confirmed that all the test cases performed as expected.
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.

o

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer submitted additional information on exclusions and meaningful differences.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Is it reasonable to assume that the impact of exclusions will be similar for the eMeasure and the chart-abstracted measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer submitted additional information on exclusions and meaningful differences.

Question for the Committee:

- *For meaningful differences in the chart-abstracted measure the developer refers to the performance scores over time. Does the SC think the eMeasure will demonstrate similar results to the chart-abstracted measure which show little to no room for improvement?*

2b6. Comparability of data sources/methods: Not Applicable

2b7. Missing Data

- The developer describes the handling of missing data but does not provide information on the frequency of missing data or potential impact on measure results due to limited testing abilities using the Bonnie tool.

Guidance from Validity Algorithm: Specifications consistent with evidence (Box 1) → threats to validity empirically assessed to some extent (Box 2)→ empirical testing of data elements using BONNIE tool (Box 3,10))→Moderate

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:

**specifications consistent

**Validity high- testing EHR derived data from paper record

**Specifications are consistent with the evidence

2a2. Reliability Testing

Comments:

**No

**Yes

**Validity testing supported the measure.

2b2. Validity Testing

Comments:

**yes

**Not enough actual data to draw a conclusion-but no reason to assume threat to validity

**Unknown

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments:
**Low reliability testing only data element level
**Not yet- no evidence that reliability will be compromised with Emeasure compared to paper chart extraction. Reliability should increase.
**Reliability was tested and indicate that the measure is repeatable

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is EHR; all data elements are in defined fields in EHR.
- The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible.
- This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment.
- Per developer, they were unable to directly assess data accuracy of the data elements in an EHR system. Workflow, or the degree to which the data elements are captured during the course of care was also not assessed.

Questions for the Committee:

- *Is it likely the required data elements are routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Concerned that developer was unable to directly assess data accuracy
**No concerns about feasibility of collecting required data elements. Discharge on appropriate antithrombotic therapy part of the "Get with the Guidelines" standards for all certified stroke programs.
**Required data elements are readily generated and the measure has been in use.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- CMS Hospital Inpatient Quality Reporting Program (IQR)/EHR Incentive Program: The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. The EHR Incentive Program provides incentive payments to promote the adoption and meaningful use of interoperable HIT and qualified EHRs.
- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results:

- The developer states that while the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are currently available.
- In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs, one of which is this measure. Data will be accepted through February 28th, 2017.

Unexpected findings (positive or negative) during implementation:

- Developer did not identify unexpected findings during implementation.

Potential harms:

- Developer did not identify any unintended consequences related to this measure.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Not publicly reported but in CMS accountability program with data accepted through Feb 2017. Usability seems promising/beneficial

**NA

**Measure has been in regular use

Criterion 5: Related and Competing Measures

Related or competing measures

- 0068 : Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet
- 0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy
- 0435 : STK 02: Discharged on Antithrombotic Therapy
- 0438 : STK 05: Antithrombotic Therapy By End of Hospital Day Two
- 0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy – American Academy of Neurology – not NQF endorsed

Harmonization :

- Measure 0438, Antithrombotic Therapy By End of Hospital Day 2, is the fifth (STK-5) measure in The Joint

Commission stroke core measure set and also targets the ischemic stroke population; however, the timeframe for antithrombotic administration is different in this measure. All common data elements for these measures are completely harmonized.

- NQF#0435: STK 02: Discharged on Antithrombotic Therapy: The measures are completely harmonized to the extent possible, given the fact that the data source for #0435 is the paper medical record, and the data source for #2832 is the electronic health record.
- Measure 0068 is a physician level measure that encompasses a different target population, specifically patients with ischemic vascular disease who were discharged alive for acute myocardial infarction (AMI), coronary artery bypass graft (CABG) or percutaneous coronary interventions (PCI).

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NQF #: 0435 NQF Project: Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The focus of the measure is to prevent a secondary stroke following a new ischemic stroke. Antithrombotic prescribed at discharge >> secondary stroke prevention >>decreased morbidity and mortality.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
This measure is consistent with the guidelines recommended by the American Heart Association/American Stroke Association for the prevention of stroke in patients with stroke or transient ischemic attack. The Antiplatelet Trialists' Collaboration (1994) and later Antithrombotic Trialists' Collaboration (2002) have confirmed that the benefit of long-term antithrombotic therapy to patients with ischemic stroke is well established.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: In 2002, the Antithrombotic Trialists' Collaboration published a systematic overview of the relevant literature supporting antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. The collaboration reviewed 287 studies involving 135,000 patients in comparisons of antiplatelet therapy versus control and 77,000 patients in comparisons of different antiplatelet regimens. Twenty-one of the 287 trials were randomized control trials (n=18,270) involving patients with a history of stroke or transient ischemic attack allocated to a mean duration of 29 months of antiplatelet therapy. Since the previous meta-analysis (1994; 10,255 patients in 18 trials), the amount of information available on the effects of long-term antithrombotic therapy among patients with a history of stroke or transient ischemic attack has substantially increased, primarily due to the results of the second European Stroke Prevention Study (n-6602).

Relevant trials were identified through electronic database searches (MEDLINE, Embase, Derwent, Scisearch, and Biosis), searching the trials registers of the Cochrane Stroke and Peripheral Vascular Disease Groups; manual searching of journals, abstracts, and proceedings of meetings; scrutinizing the reference lists of trials and review articles; and professional inquiry, including colleagues and representatives of pharmaceutical companies.

Trials available by September 1997 that compared an antiplatelet regimen with a control or one antiplatelet regimen with another among patients considered to be high annual risk (> 3%/year) of vascular events because of evidence of pre-existing disease were included in the meta-analysis. Only those trials believed to have used a randomization method that precluded prior knowledge of the next treatment allocated and contained two randomized groups that differed only with respect to the antiplatet comparison of interest were selected. Furthermore, trials of oral antiplatet regimens were eligible only if they had assessed more than one day of treatment. An antiplatelet drug was defined as one whose primary effect on the vascular system inhibits platelet adhesion, platelet aggregation, or both. Details about the method of randomization, blinding of treatment allocation, scheduled duration of treatment, and, if, different,

scheduled duration of follow-up were requested of all trial coordinators.

The primary measure of outcome was a "serious vascular event" (i.e., non-fatal myocardial infarction, non-fatal stroke, or death from a vascular cause including deaths from an unknown cause). Deaths were divided into those with a vascular or non-vascular cause. Strokes were subdivided into intracranial hemorrhages (including intracerebral, subdural, subarachnoid, and extradural hemorrhages), ischemic strokes, and strokes of unknown etiology.

In addition to the meta-analysis from the Antithrombotic Trialists' Collaboration, the following trials relevant to antithrombotic therapy for noncardioembolic stroke were noted in the literature. The landmark trials primarily focused on aspirin and included: the International Stroke Trial Collaborative Group (1994); the Chinese Acute Stroke Trial Collaborative Group (1997); the European Stroke Prevention Study Group 2 (1997); the Canadian Cooperative Study Group (1978); the UK-TIA Study Group (1991); the Dutch TIA Study Group (1991); and the SALT Collaborative Group (1991). Three RCTs evaluated ticlopidine: the Canadian American Ticlopidine Study (1989), the Ticlopidine Aspirin Stroke Study (1989), and the African American Antiplatelet Stroke Prevention Study (2003). Clopidogrel was compared to aspirin alone in the Clopidogrel versus Aspirin in Patients at Risk of Ischemic Events (CAPRIE) trial (1996). The Management of Atherothrombosis with Clopidogrel in High-Risk Patients with Transient Ischemic Attacks or Ischemic Strokes (MATCH) trial (2004) evaluated the effectiveness of clopidogrel 75mg and aspirin 75 mg, compared with clopidogrel 75mg alone for stroke prevention. Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization, Management, and Avoidance (CHARISMA) trial (2006) and the Fast Assessment of Stroke and Transient ischemic attack to prevent Early Recurrence (FASTER) trial (2007) evaluated clopigogrel and aspirin in combination with aspirin alone. Four large RCTs evaluated the effectiveness of dipyridamole and aspirin among patients with stroke or transient ischemic attack: European Stroke Prevention Studies (ESPS-1 1987 and ESPS-2 1997); the European Australian Stroke Prevention in Reversible Ischemic Stroke (ESPRIT) trial (2006); and the Prevention Regimen for Effectively Avoiding Second Strokes (PRoFESS) trial (2008).

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting antithrombotic therapy for secondary stroke prevention is high. Antithrombotic therapy post ischemic stroke is protective for high risk patients, such as, those with a history of prior stroke or vascular event. The body of evidence established by the Antithrombotic Trialists' Collaboration confirmed that prolonged antithrombotic therapy reduces the relative risk of stroke, myocardial infarction or death by approximately 22%. Risk of gastrointestinal hemorrhage or other major hemorrhage is the primary concern with ongoing therapy, although the studies have demonstrated that the risk is low and absolute benefits significantly outweigh the risks.

There was no appreciable evidence that either higher aspirin dose or any other antiplatelet regimen was more effective than low dose aspirin (75 mg to 150 mg) in preventing vascular events. Additionally, the body of evidence did not establish the optimal duration of treatment. The majority of studies reviewed lasted one to three years. The evidence seems to support that longer treatment might be more effective due to significant further benefit noted ($P<0.00001$) between year one and year three for most studies. Adding a second antithrombotic to aspirin may produce additional benefit; however, the evidence is inconclusive in this respect and more research recommended.

In reviewing the literature, there have been more studies conducted than included in the Antithombotic Trialists' meta-analysis. Some studies were not included in the meta-analysis due to the method of randomization chosen or small sample sizes. Antithrombotic medications other than aspirin have not been studied as extensively as aspirin.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of evidence consistently supports that prolonged antithrombotic therapy reduces the risk of death in patients with a prior history of ischemic stroke. Based on these findings, antithrombotic therapy should be prescribed at discharge following acute ischemic stroke, unless there are contraindications to therapy.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
There is strong evidence on the immediate hazards and net benefits of long-term antithrombotic therapy for the prevention of stroke recurrence. The Antithrombotic Trialists' Collaboration meta-analysis found a marginally significant ($P=0.04$) reduction in vascular mortality (7 deaths per 1,000 patients with a history of stroke or transient ischemic attack treated with antithrombotic therapy); however, the reduction in non-fatal vascular events and in all cause mortality was highly significant ($P=0.002$; 15 deaths per 1,000 patients

treated). These benefits exceeded the risk of bleeding (1-2 additional extracranial bleeds per 1,000 patients/year). On average, antithrombotic medications reduce the relative risk of stroke, MI, or death, but important differences exist between the various drugs in this category which directly impacts therapeutic selection.

Considering the cost of aspirin compared to the mean lifetime cost of ischemic stroke, estimated at $140,048 per person, antithrombotic therapy as a secondary stroke prevention strategy is clearly cost-effective. Gaspoz and colleagues (2002) concluded that the extension of aspirin therapy from the current levels of use to all eligible patients for 25 years would have an estimated cost-effectiveness ratio of about $11,000 per quality-adjusted year of life gained. The addition of clopidogrel for the 5 percent of patients who are ineligible for aspirin would cost about $31,000 per quality-adjusted year of life gained. Clopidogrel alone in all patients or in routine combination with aspirin had an incremental cost of more than $130,000 per quality-adjusted year of life gained and remained financially unattractive across a wide range of assumptions. However, clopidogrel alone or in combination with aspirin would cost less than $50,000 per quality-adjusted year of life gained if its price were reduced by 70 to 82 percent, to $1.00 and $0.60 per day, respectively.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: American Heart Association. During our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, and provided information that was accessible and met the requirements set out in this measure maintenance form.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.13 Grade Assigned to the Body of Evidence: Class I, Level of Evidence A

1c.14 Summary of Controversy/Contradictory Evidence: The benefit of prolonged antithrombotic therapy post ischemic stroke is undisputed. No position against long-term antithrombotic therapy was noted in the literature. However, there is still not enough evidence to answer some questions. More randomized trials are needed to determine the optimal therapeutic agent(s) and regimen(s), as well as, the optimal duration of therapy.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and thrombolytic therapy for ischemic stroke: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest. 2004; 126;483S-512S.
- Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy, I: prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. BMJ. 1994;308:81–106.
- Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomized trials of anti-platelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients [published correction appears in BMJ. 2002;324:141]. BMJ. 2002;324:71– 86.
- Bhatt DL, Fox KAA, Hacke W, Berger PA, Black HR, Boden WE, Cacoub P, Cohen EA, Creager MA, Easton J, Flather M, Haffner S, Hamm C, Hankey G, Johnston S, Mak K, Mas J, Montalescot G, Pearson T, Steg P, Steinhubl S, Weber M, Brennan D, Fabry-Ribaudo L, Booth J, Topol E; CHARISMA investigators. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. NEJM. 2006;354:1706 –1717.
- CAPRIE Steering Committee. A randomized, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). Lancet. 1996;348:1329 –1339.
- Diener H-C, Bogousslavsky J, Brass LM, Cimminiello C, Csiba L, Kaste M, Leys D, Matias-Guiu J, Rupprecht H-J; on behalf of the MATCH investigators. Aspirin and clopidogrel compared with clopidogrel alone after ischaemic stroke or transient ischaemic attack in high-risk patients (MATCH): randomized, double-blind, placebo-controlled trial. Lancet.

2004;364:331–337.
• Gaspoz J, Coxson PG, Goldman PA, Williams LW, Kuntz KM, Hunink MGM, Goldman L. Cost effectiveness of aspirin, clopidogrel, or both for secondary prevention of coronary heart disease. NEJM. 2002;346:1800-1806.
• Gorelick PB, Richardson D, Kelly M, Ruland S, Hung E, Harris Y, Kittner S, Leurgans S; for the African American Antiplatelet Stroke Prevention Study (AAASPS) Investigators. Aspirin and ticlopidine for prevention of recurrent stroke in black patients. JAMA. 2003;289: 2947–2957.
• Gent M, Easton JD, Hachinski VC, Panak E, Sicurella J, Blakely JA, Ellis DJ, Harbison JW, Roberts RS, Turpie AGG. The Canadian American Ticlopidine Study (CATS) in Thromboembolic Stroke. Lancet. 1989;1215-1220.
• Guyatt GH, Akl EA, Crowther M, Gutterman DD, Sch?nemann HJ, and for the American College of Chest Physicians Antithrombotic Therapy and Prevention of Thrombosis Panel. Executive summary: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141: 34S.
• Guyatt G, Schunëmann H, Cook D, Jaeschke R, Pauker S, Bucher H. Grades of recommendation for antithrombotic agents. Chest. 2001;119:3S-7S.
• Hass WK, Easton JD, Adams HP, Pryse-Phillips W, Molony BA, Anderson S, Kamm B; for the Ticlopidine Aspirin Stroke Study Group. A randomized trial comparing ticlopidine hydrochloride with aspirin for the prevention of stroke in high-risk patients. N Engl J Med. 1989;321: 501–507.
• Johnson ES, Lanes SF, Wentworth CE, Satterfield MH, Abebe BL, Dicker LW. A meta-regression analysis of the dose-response effect of aspirin on stroke. Arch Intern Med. 1999;159:1248 –1253.
• Kennedy J, Hill MD, Ryckborst KJ, Eliasziw M, Demchuk AM, Buchan AM. Fast assessment of stroke and transient ischaemic attack to prevent early recurrence (FASTER): a randomised controlled pilot trial. Lancet Neurol. 2007;6:961–969.
• Sacco RL, Diener H-C, Yusuf S, Cotton D, Ounpuu S, Lawton WA, Palesch Y, Martin RH, Albers GW, Bath P, Bornstein N, Chan BP, Chen ST, Cunha L, Dahlo¨f B, De Keyser J, Donnan GA, Estol C, Gorelick P, Gu V, Hermansson K, Hilbrich L, Kaste M, Lu C, Machnig T, Pais P, Roberts R, Skvortsova V, Teal P, Toni D, Vandermaelen C, Voigt T, Weber M, Yoon BW; PRoFESS Study Group. Aspirin and extended release dipyridamole versus clopidogrel for recurrent stroke. N Engl J Med. 2008;359:1238 –1251.
• Sandercock P, Collins R, Counsell C, Farrell B, Peto R, Slattery J, Warlow C, International Stroke Trial Collaborative Group: The International Stroke Trial (IST): A randomised trial of aspirin, subcutaneous heparin, both, or neither among 19,435 patients with acute ischaemic stroke. Lancet. 1997;349:1569-1581.
• The Canadian Cooperative Study Group. A randomized trial of aspirin and sulfinpyrazone in threatened stroke. N Engl J Med. 1978;299:53–59.
• The Dutch TIA Trial Study Group. A comparison of two doses of aspirin (30 mg vs. 283 mg a day) in patients after a transient ischemic attack or minor ischemic stroke. N Engl J Med. 1991;325:1261–1266.
• The ESPRIT Study Group. Aspirin plus dipyridamole versus aspirin alone after cerebral ischaemia of arterial origin (ESPRIT): randomised controlled trial. Lancet. 2006;367:1665–1673.
• The ESPS Group. The European Stroke Prevention Study (ESPS): principal end-points. Lancet. 1987:2:1351–1354.
• The SALT Collaborative Group. Swedish Aspirin Low-dose Trial (SALT) of 75 mg aspirin as secondary prophylaxis after cerebrovascular ischaemic events. Lancet. 1991;338:1345–1349.
• UK-TIA Study Group. The United Kingdom Transient Ischaemic Attack (UK-TIA) aspirin trial: final results. J Neurol Neurosurg Psychiatry. 1991;54:1044 –1054.

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)*:
2010 AHA/ASA Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack, page 249.

Class I, Level of Evidence A

For patients with noncardioembolic ischemic stroke or TIA, the use of antiplatelet agents rather than oral anticoagulation is recommended to reduce the risk of recurrent stroke or another cardiovascular event.

Aspirin (50 mg/d to 325 mg/d) monotherapy (Class I, Level of Evidence A), the combination of aspirn 25 mg and extended-release dipyridamole 200 mg twice daily (Class I, Level of Evidence B), and clopidogrel 75 mg monotherapy daily (Class IIa, Level of Evidence B), are all acceptable options for initial therapy. The selection of an antiplatelet agent should be individualized on the basis of patient risk factor profiles, cost, tolerance, and other clinical characteristics.

1c.17 Clinical Practice Guideline Citation: Furie KL, Kasner SE, Adams RJ, Albers GW, et al. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2011;42:246-49.

1c.18 National Guideline Clearinghouse or other URL: http://stroke.ahajournals.org/content/42/1/227.full.pdf

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: This guideline was approved by the American Heart Association Science Advisory and Coordinating Committee on July 13, 20010. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence

from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.23 Grade Assigned to the Recommendation: Class I, Level of Evidence A

1c.24 Rationale for Using this Guideline Over Others: The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: High 1c.26 Quality: High1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0435_Evidence_MSF5.0_Data-635827722515582215.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. The effectiveness of antithrombotic agents in reducing stroke mortality, stroke-related morbidity and recurrence rates has been studied in several large clinical trials. Studies have shown that antithrombotic therapy prescribed at discharge following acute ischemic stroke reduces stroke mortality and morbidity.

Healthcare organizations that track antithrombotic therapy prescribed at discharge for internal quality improvement purposes have seen an improvement in the measure rate over time. The chart-abstracted measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to which will also promote improvements in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This measure is a legacy eCQM that is currently included in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data for the electronic version of the measure are yet available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs. This measure is one of the 28.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
The use of antithrombotic therapy after ischemic stroke is recommended for secondary stroke prevention. Earlier evidence supported substantial underutilization of this therapy, particularly among the very elderly, those admitted from skilled nursing facilities, and patients with functional dependence (Lichtman JH, 2011). An analysis of Medicare data for 31,554 non-terminally ill patients with ischemic stroke randomly selected from the Medicare Health Care Quality Improvement Program's National Stroke Project (1998 to 1999, 2000 to 2001) demonstrated that only 74.2% were discharged on an antithrombotic medication. Treatment rates decreased with age and were lowest for patients' age 85-years and older. Later data demonstrates that this previous performance gap of 25% has narrowed significantly.

In October, 2009, The Joint Commission added the chart-abstracted stroke (STK) measure set as a new core measure option to meet performance measure requirements for Joint Commission hospital accreditation purposes. Below is the specified level of analysis for STK-2 beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 2034 denominator cases; 1995 numerator cases; 49 hospitals; 0.98083 national aggregate rate; 0.98731 mean of hospital rates; 0.2575 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.98601 25th percentile rate/lower quartile; and, 0.94737 10th percentile rate.

CY 2010: 19,889 denominator cases; 19,622 numerator cases; 137 hospitals; 0.98658 national aggregate rate; 0.97644 mean of hospital rates; 0.05765 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99711 50th percentile rate/median rate; 0.98276 25th percentile rate/lower quartile; and, 0.93436 10th percentile rate.

CY 2011: 24,053 denominator cases; 23,812 numerator cases; 157 hospitals; 0.98998 national aggregate rate; 0.97173 mean of hospital rates; 0.10346 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99803 50th percentile rate/median rate; 0.98529 25th percentile rate/lower quartile; and, 0.94915 10th percentile rate.

CY 2012: 24,420 denominator cases; 24,217 numerator cases; 158 hospitals; 0.99169 national aggregate rate; 0.98674 mean of hospital rates; 0.04061 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.98913 25th percentile rate/lower quartile; and, 0.97059 10th percentile rate.

CY 2013: 37,661 denominator cases; 37.363 numerator cases; 262 hospitals; 0.99209 national aggregate rate; 0.9885 mean of hospital rates; 0.02736 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile

rate/median rate; 0.98913 25th percentile rate/lower quartile; and, 0.96667 10th percentile rate.

CY 2014: 180,048 denominator cases; 178,945 numerator cases; 1299 hospitals; 0.99387 national aggregate rate; 0.98632 mean of hospital rates; 0.06957 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.99254 25th percentile rate/lower quartile; and, 0.97674 10th percentile rate.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al., 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

In the national REasons for Geographic and Racial Differences in Stroke (REGARDS) cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years of age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age : RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates

in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer,et al., 2010).

In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).

Potential disparities in secondary pharmacological prevention for recurrent stroke have been identified. A study using the uniform data system for medical rehabilitation data from nursing homes in five states found that blacks or African Americans were significantly less likely to receive antithrombotic therapy (Ottenbacher KJ, et al, 2001). In a model adjusting for age, sex, physical function, and comorbidities, blacks or African Americans proved to be 80% as likely as whites to receive antithrombotic therapy.

Based on a national cross-sectional study of nursing home residents, the use of any antithrombotic (anticoagulant or antiplatelet) therapy was lower for Hispanics (45.5%), non-Hispanic blacks (49.4%), and Asian/Pacific Islanders (38.9%) than for whites (54.3%), although higher among a subsample of Native Americans (58.0%). Aspirin use was comparable across ethnic groups, and actually higher for Native American, blacks or African Americans, and Hispanic than for whites. In contrast, oral anticoagulation (warfarin) use was significantly lower for ethnic minority groups (25.4%-31.7%), except Native Americans whose proportion was only slightly lower (36.4%) than whites (39.6%)(Christian JB, et al. 2003).

Since the last endorsement date, Schwamm and colleagues (2010) found significant and important differences in quality of care related to antithrombotics at discharge when multivariate models were constructed adjusting for patient-level characteristics only. Race/ethnicity Black versus White: unadjusted OR 0.86 [95% CI 0.83-0.90]; adjusted for patient characteristics OR 0.81 [95% CI 0.78-0.85]; adjusted for patient and hospital characteristics OR 0.88 [95% CI 0.84-0.92]. Race/ethnicity Hispanic versus White: unadjusted OR 0.85 [95% CI 0.75-0.90]; adjusted for patient characteristics OR 0.80 [95% CI 0.75-0.86]; adjusted for patient and hospital characteristics OR 0.90 [95% CI 0.82-0.97]. Total N=46,515; All n 94.98%; White 95.15%; Black 94.41%; Hispanic 94.31%.

A more recently published study (Qian F, et al, 2013) from Get With The Guidelines (GWTG) also noted racial and ethnic disparities for antithrombotics at discharge. Using patient data (n=200,900) from the American Heart Association/American Stroke Association GWTG-Stroke program from April 2003 through December 2008, the following performance measure rates were reported for discharged antithrombotic: non-Hispanic White (n=170,694) 95.3%; non-Hispanic Black (n=20,514) 94.3%; Hispanic (n=6632) 94.5%; and non-Hispanic Asian American (n=3060) 94.5%. Data from the Paul Coverdell National Acute Stroke Registry (PCNASR) were similar for antithrombotic therapy at discharge (n=58,823) White 98.4%; Other Race 98.3% (Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, 2014).

• Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, Yan T, Escarce JJ. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-8.
• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Christian JB, Lapane KL, Toppa RS. Racial disparities in receipt of secondary stroke prevention agents among nursing home residents. Stroke. 2003;34:2693-2697.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, Griffith P, Gorelick PB, Howard G, Leira EC, Morgenstern LB, Ovbiagele B, Peterson E, Rosamond W, Trimble B, Valderrama AL, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• Gillum RF, Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 2011;42(11):3294-6.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-75.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.

• Jacobs BS, Birbeck G, Mullard AJ, Hickenbottom S, Kothari R, Roberts S, Reeves MJ. Quality of hospital care in African Americans and white patients with ischemic stroke and TIA. Neurology.2006;66:809-14.
• Johnston SC, Fung LH, Gillum LA, Smith WS, Brass LM, Lichtman JH, Brown AN. Utilization of intravenous tissue-type plasminogen activator for ischemic stroke at academic medical centers.: the influence of ethnicity. Stroke. 2001; 32:1061-68.
• Karve S, Balkrishnan R, Seiber E Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
•Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
•Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
•Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, Kerber KA, Skolarus LE, Meurer WJ, Burke JF, Adelman EE, Baek J, Lisabeth LD. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol. 2013;74:778–785.
•Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Ottenbacher KJ, Smith PM, Illig SB, Fiedler RC, Gonzales V, Granger CV. Characteristics of persons rehospitalized after stroke rehabilitation. Arch Phys Med Rehabil. 2001;82:1367-1374.
•Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
•Schwamm LH, Syed FA, Reeves MJ, Smith EE, Saver JL, Messe S, Bhatt DL, Grau-Sepulveda MV, Peterson ED, Fonarow GC. Temporal trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With the Guidelines-Stroke hospitals. Circ Cardiovasc Quality Outcomes. 2013;6:543-549.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States remains attributable to stroke. More women than men die of stroke each year. Women accounted for almost 60% of US stroke deaths (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, et al., 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 ($P<0.05$) (US Burden of Disease Collaborators, 2013) . Among Medicare patients discharged from the hospital after stroke,

˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

Antithrombotic agents significantly reduce the incidence of a recurrent vascular event after a stroke. The effectiveness of antithrombotic agents in reducing stroke mortality, stroke-related morbidity and recurrence rates has been studied in several large clinical trials (Kernan WN, et al, 2014; Sandercock P, et al, 2014)). While the use of these agents for patients with acute ischemic stroke and transient ischemic attacks continues to be the subject of study, substantial evidence is available from completed studies. Data at this time suggest that antithrombotic therapy should be prescribed at discharge following acute ischemic stroke to reduce stroke mortality and morbidity as long as no contraindications exist.

1c.4. Citations for data demonstrating high priority provided in 1a.3

• Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A Guideline From the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1655-1711.
• Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and Thrombolytic Therapy for Ischemic Stroke. Chest Vol. 119, 2001: 300-320.
• Brott TG, Clark WM, Grotta JC, et al. Stroke the first hours. Guidelines for acute treatment. Consensus Statement. National Stroke Association. 2000.
• Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Chen ZM, Sandercock P, Pan HC, et al. Indications for early aspirin use in acute ischemic stroke: a combined analysis of 40,000 randomized patients from the Chinese acute stroke trial and the international stroke trial. On behalf of the CAST and IST collaborative groups, Stroke 2000;31:1240-1249.
• Coull BM, Williams LS, Goldstein LB, et al. Anticoagulants and Antiplatelet Agents in Acute Ischemic Stroke. Report of the Joint Stroke Guideline Development Committee of the American Academy of Neurology and the American Stroke Association (a Division of the American Heart Association) Stroke. 2002;33:1934 -1942.
• Furie KL, Kasner SE, Adams RJ, Albers GW, et al. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2011;42:241-43.
• George M., Xin Tong, McGruder H., Yoon P., Rosamond W., Winquist A., Hinchey J., Wall H., Pandey D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Guideline on the Use of Aspirin as Secondary Prophylaxis for Vascular Disease in Primary Care, Centre for Health Services Research University of Newcastle upon Tyne, & Centre for Health Economics of York, 1998.
•Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, Fang MC, Fisher M, Furie KL, Heck DV, Johnston SC, Kasner SE, Kittner SJ, mitchell PH, Rich MW, Richardson D, Schwamm LH, Wilson JA. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2014;45:241-2160-2236.
•Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e-78-82, e-119-127.
•Sandercock P, Counsell C, Tseng MC, Cecconi E. Oral antiplatelet therapy for acute ischaemic stroke. The Cochrane Database of Systematic Reviews. 2014;3:CD000029.

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input was obtained.)*

Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/ecqm_library.html

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is an eMeasure **Attachment:** STK2_MAT.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** DischargedonAntithromboticThe_v4_Mon_Apr_06_11.08.32_CDT_2015.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Changes have been made to the eCQM specifications in order to reflect revisions to the chart abstracted measure from which this measure is derived.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients prescribed antithrombotic therapy at hospital discharge.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Antithrombotic Therapy Prescribed at Discharge

- Antithrombotic Therapy is represented with the QDM datatype and value set of Medication, Discharge: Antithrombotic Therapy (OID: 2.16.840.1.113883.3.117.1.7.1.201)

Non-Elective Inpatient Encounter

- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Patients with a principal diagnosis of ischemic stroke.

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Principal Diagnosis of Ischemic Stroke

- Ischemic Stroke is represented with the QDM datatype and value set of Diagnosis, Active: Ischemic Stroke (OID: 2.16.840.1.113883.3.117.1.7.1.247)
- Ordinality: Principal (OID: 2.16.840.1.113883.3.117.1.7.1.14)

Non-Elective Inpatient Encounter

- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Denominator Exclusions:
Patients with comfort measures documented.
Patients admitted for elective carotid intervention. This exclusion is implicitly modeled by only including non-elective hospitalizations.
Patients discharged to another hospital
Patients who left against medical advice
Patients who expired
Patients discharged to home for hospice care
Patients discharged to a health care facility for hospice care

Denominator Exceptions:
Patients with a documented reason for not prescribing antithrombotic therapy at discharge.

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Denominator Exclusion Data Elements:

Non-Elective Inpatient Encounter

- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

Discharge Status (modeled as Attributes of the above Non-Elective Inpatient Encounter)

- Discharge status: Left Against Medical Advice (OID: 2.16.840.1.113883.3.117.1.7.1.308)
- Discharge status: Patient Expired (OID: 2.16.840.1.113883.3.117.1.7.1.309)
- Discharge status: Discharge To Acute Care Facility (OID: 2.16.840.1.113883.3.117.1.7.1.87)
- Discharge status: Discharged to Home for Hospice Care (OID: 2.16.840.1.113883.3.117.1.7.1.209)
- Discharge status: Discharged to Health Care Facility for Hospice Care (OID: 2.16.840.1.113883.3.117.1.7.1.207)

Comfort Measures

- Comfort Measures are represented with the QDM datatypes and value set of:
- Intervention, Order: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)
- Intervention, Performed: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)

Emergency Department Visit
- Emergency Department Visit is represented with the QDM datatype and value set of Encounter, Performed: Emergency Department Visit (OID: 2.16.840.1.113883.3.117.1.7.1.292)

Denominator Exceptions Data Elements:

Reasons for Not Ordering Antithrombotic Medication
- Antithrombotic Ingredient Specific Medication is represented with the QDM datatype and value set of Medication, Discharge: Antithrombotic ingredient specific (OID: 2.16.840.1.113762.1.4.1021.8)
- Medical Reason is represented with the QDM datatype and value set of Medication, Discharge not done: Medical Reason (OID: 2.16.840.1.113883.3.117.1.7.1.473)
- Patient Refusal is represented with the QDM datatype and value set of Medication, Discharge not done: Patient Refusal (OID: 2.16.840.1.113883.3.117.1.7.1.93)

Emergency Department Visit
- Emergency Department Visit is represented with the QDM datatype and value set of Encounter, Performed: Emergency Department Visit (OID: 2.16.840.1.113883.3.117.1.7.1.292)

Non-Elective Inpatient Encounter
- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable, the measure is not stratified.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
See attached HQMF file.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
eMeasures are calculated using only the structured data collected in certified EHR technology (CEHRT). Data not present in the structured field from which the measure draws will not be included in the measure calculation.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Pharmacy

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Hospitals report EHR data using Certified Electronic Health Record Technology (CEHRT), and by submitting Quality Reporting Document Architecture Category 1 (QRDA-1).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF0435_CMS104v4_STK2_Bonnie_Testing.xlsx,STK2_eCQM_testing_attachment-635907892750810521.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2832
Measure Title: STK02: Discharged on Antithrombotic Therapy
Date of Submission: 1/14/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	X Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the sub criteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of

sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒other: Bonnie Test Cases	☒ other: Bonnie Test Cases (see question 1.2)

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

This measure is a legacy eCQM (an NQF endorsed measure that has been re-specified into an eMeasure and is currently used in a federal quality program/programs). In accordance with the modified NQF requirements for testing of eMeasures, the Bonnie testing tool was used to simulate a testing environment where measure specifications and HQMF output are tested against synthetic test data. Measure developers rely on the results in Bonnie to confirm whether the measure logic is performing as expected. Reference the eCQI Resource Center website (https://ecqi.healthit.gov/ecqm-tools/tool-library/bonnie) or the Bonnie testing tool website (https://bonnie.healthit.gov/) for more information about Bonnie functionality and its role in measure development. Please also reference the Bonnie testing worksheet attachment for detailed Bonnie test cases and testing results for this measure.

While Bonnie does consume HQMF specifications, we are not able to test the measure with data from HQMF implemented in EHRs.

1.3. What are the dates of the data used in testing? The measure specifications were tested in the Bonnie testing environment which mimics the year 2012.

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☒ other: Bonnie Test Cases

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Not applicable.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

22 unique synthetic patient records were created in the BONNIE testing system for this measure. Cases were used to test the validity of each data element and timing relationship in the measure. Patient characteristics such as age, diagnosis, and medications were pre-determined to provide a variety of scenarios that adequately test patients passing each data element and failing each data element. Data included in cases and tested for this measure included diagnoses, medications, discharge statuses, patient orders, age, length of stay, and ED and inpatient encounters.

All 22 cases passed or failed as expected based on the data included in the case, confirming the measure logic is accurate and valid.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Not applicable. Bonnie test patients were developed for use across the different aspects of testing.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Not applicable.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*May be one or both levels*)

X **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

The chart-abstracted version of this measure has been in national use since the 4th quarter of 2009. At the time the chart-abstracted measure was originally tested, extensive tests of measure reliability were conducted. At present, no performance data for the electronic version of the measure are currently available. Below is the reliability testing summary for NQF #0435 STK02: Discharged on Antithrombotic Therapy, from which this measure is derived.

At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with monthly data transmission to The Joint Commission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of approximately 900 patient records identified at 30 pilot sites over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rates of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify potential measure enhancements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures.) The following is a list of the major tests conducted on the submitted data for this measure:

- Transmission of complete data
- Usage of data received (to understand if the healthcare organization provides the relevant service to treat the relevant population)
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity
- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors is as follows:

- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data element agreement rates reported to The Joint Commission for the time period of one year (4Q2010 – 3Q2011) have shown an overall agreement rate of 97.61%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-2.

Data Elements	Total 'n' numerator	Total 'n' denominator	Rate
Antithrombotic Therapy Prescribed at Discharge	442	460	96.1%
Clinical Trial	712	714	99.7%
Comfort Measures Only	709	714	99.3%
Elective Carotid Intervention	711	714	99.6%
Reason for Not Prescribing Antithrombotic Therapy at Discharge	10	11	91.0%

These agreement rates are considered to be well within acceptable levels.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Agreement rates for individual data elements tested for the chart-abstracted version of this measure were within acceptable levels. Once data are available for analysis, it is expected that reliability tests of the eCQM version of this measure will yield similar results.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

- ☒ **Critical data elements** (*data element validity must address ALL critical data elements*)
- ☒ **Performance measure score**
  - ☐ **Empirical validity testing**
  - ☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The Bonnie testing tool and environment were used to test the measure logic and value sets. Each data element and logic statement was tested to confirm actual results met expectations. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure. An example of negative testing would be to include test cases with pediatric ages to ensure that pediatric patients are not included in the measure.

Denominator test cases positively test to ensure patients with principal diagnosis of ischemic stroke appropriately fall into the measure. We have created negative test cases, testing to ensure patients with a different principal diagnosis do not actually make it into the denominator.

Numerator test cases positively test to ensure patients receiving antithrombotic therapy at discharge pass the measure. Negative test cases ensure that a patient who does not receive antithrombotic therapy in the specified time frame do not pass the measure.

Denominator exclusion and exception test cases for this measure ensure that patients are properly removed from the denominator if they have comfort measures, specific discharge statuses other than discharged home, and medical reasons or patient refusal for not getting antithrombotic therapy. Negative test cases are also run, an example of this would be comfort measures that are ordered but not in the specified time frame. Testing confirmed patients meeting the exclusion and exception criteria are removed from the measure appropriately.

A review of the measure specifications was also conducted to confirm the logic was properly expressed within the current version of the QDM and confirmed the logic matches the clinical intent of the measure, as stated in the measure header. This is done through use of a logic checklist to facilitate review of the measure logic, according to several checks, a few examples of which are included below:

- Is the intent of the measure described in the measure description articulated/ captured in the measure logic?
- Do the logic elements map to definitions in the measure narrative, data dictionary or supporting reference documentation?
- Do the populations in the narrative align with the populations defined in the logic?
- Are the mathematic inequalities reflective of the measure intent and represent the intended populations (for example: when intended the inequality represents less than rather than less and or equal to)?

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Bonnie results provide coverage and passing rates. This measure reached 100% coverage, confirming there is a test case for each pathway of logic. This measure also has a 100% passing rate, confirming all test cases performed as expected.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

This measure performed as expected in Bonnie.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Exclusions in the eCQM align with the chart-based version of the measure, and are clinically necessary for the interpretation of the measure. As noted previously, the chart-abstracted version of this measure has been in national use since the 4th quarter of 2009, and no data are available for the eCQM version of the measure. The below analysis addresses exclusions testing performed for the chart-abstracted version of the measure from which this measure is derived.

The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patient who have a Length of Stay greater than 120 days
- Patients with Comfort Measures
- Patient enrolled in a Clinical Trial
- Patients admitted for Elective Carotid Intervention
- Patients with a documented Reason For Not Prescribing Antithrombotic Therapy at Discharge
- Patients discharged to another hospital
- Patients who left against medical advice
- Patients who expired
- Patients discharged to home for hospice care
- Patients discharged to a health care facility for hospice care

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are detailed below.

Exclusion: Comfort Measures Only
Overall Occurrence n = 163,225
Overall Occurrence Percentage: 10.4%
Minimum: 0.30%
10th Percentile: 5.26%
Median: 10%
90th Percentile: 15.8%
Maximum: 35.7%

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%
10th Percentile: 0.24%
Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for Elective Carotid Intervention
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.5%
90th Percentile: 25.3%
Maximum: 95.2%

Exclusion: Patients with a documented Reason For Not Prescribing Antithrombotic Therapy at Discharge
Overall Occurrence n = 19,751
Overall Occurrence Percentage: 6.27%
Minimum: 0.27%

10th Percentile 0.92%
Median: 2.82%
90th Percentile: 16.1%
Maximum: 60.6%

Exclusion: Discharge Disposition - Patients discharged to another hospital
Overall Occurrence n = 449,924
Overall Occurrence Percentage: 35.7%
Minimum: 0.787%
10th Percentile: 25%
Median: 35.4%
90th Percentile: 46%
Maximum: 76.2%

Exclusion: Discharge Disposition - Patients who left against medical advice
Overall Occurrence n = 8,396
Overall Occurrence Percentage: 0.67%
Minimum: 0.067%
10th Percentile: 0.34%
Median: 0.86%
90th Percentile: 2.4%
Maximum: 9.67%

Exclusion: Discharge Disposition - Patients who expired
Overall Occurrence n = 76,168
Overall Occurrence Percentage: 6.04%
Minimum: 0.398%
10th Percentile: 1.9%
Median: 5.08%
90th Percentile: 10.2%
Maximum: 20.1%

Exclusion: Discharge Disposition - Patients discharged to home for hospice care
Overall Occurrence n = 658,264
Overall Occurrence Percentage: 52.2%
Minimum: 6.25%
10th Percentile: 39%
Median: 51.9%
90th Percentile: 64%
Maximum: 94.3%

Exclusion: Discharge Disposition - Patients discharged to a health care facility for hospice care
Overall Occurrence n = 37,804
Overall Occurrence Percentage: 3%
Minimum: 0.169%
10th Percentile: 0.87%
Median: 3.01%
90th Percentile: 6.4%
Maximum: 23%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Analysis of these data for the chart-asbtracted progenitor of this measure indicated that all exclusions were appropriate. It is believed that results for exclusions in the eCQM will be similar when sufficient data have been received to perform such an analysis.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

- ☒ **No risk adjustment or stratification**
- ☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
- ☐ **Stratification by** Click here to enter number of categories **risk categories**
- ☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

A similar methodology will be used for the eCQMs.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

NQF#0435: STK-2 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10th Percentile
3Q2011	154	0.98351	0.07256	1	1	1	1	0.97674
2Q2011	155	0.97524	0.11930	1	1	1	1	0.96491
1Q2011	160	0.97249	0.10725	1	1	1	0.98833	0.94495
4Q2010	135	0.98048	0.06361	1	1	1	1	0.94286
3Q2010	130	0.96843	0.11476	1	1	1	0.98889	0.94291
2Q2010	122	0.97233	0.08071	1	1	1	0.97959	0.94203
1Q2010	99	0.97384	0.06745	1	1	1	0.97959	0.91667
4Q2009	49	0.98731	0.02575	1	1	1	0.98601	0.94737

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across

measured entities? (i.*e.*, *what do the results mean in terms of statistical and meaningful differences?*)

It should be noted that since data collection on this measure is completely voluntary for both The Joint Commission and PCNASR, the hospitals reporting on this measure are self-selected, and therefore, presumably have a particular interest and concern for improving stroke care. For this reason, the measure results demonstrated by this group most likely significantly overstates the rate for the population of all health care organizations.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Multiple data sources are not used for this measure.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e.*, *what do the results mean and what are the norms for the test conducted*)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Data not present in the structured field from which the measure draws will not be included in the measure calculation. In the Bonnie testing environment, missing data are tested as an expected "Fail" for that data element, and actual performance (whether or not the case fails) is compared to expectations to ensure missing

data impacts data element and overall measure calculation as expected. This is the extent of missing data analysis that can be performed in Bonnie.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

See response for question 2b7.1 above.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

The measure performed as expected in the Bonnie testing environment.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment **Attachment:** STK2_eCQM_NQF_Measure_Feasibility_Assessment_Report.docx

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Not applicable.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

Value sets are housed in the Value Set Authority Center (VSAC), which is provided by the National Library of Medicine (NLM), in coordination with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

Viewing or downloading value sets requires a free Unified Medical Language System® (UMLS) Metathesaurus License, due to usage restrictions on some of the codes included in the value sets. Individuals interested in accessing value set content can request a UMLS license at (https://uts.nlm.nih.gov/license.html)

There are no other fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program http://jointcommission.org EHR Incentive Program https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentivepr

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

- Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
- Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 hospitals (2015)

- Name of program and sponsor: EHR Incentive Program; Centers for Medicare & Medicaid Services
- Purpose: The American Recovery and Reinvestment Act of 2009 (ARRA) (Pub.L. 111–5) was enacted on February 17, 2009. Title IV of Division B of ARRA amends Titles XVIII and XIX of the Social Security Act (the Act) by establishing incentive payments to eligible professionals (EPs), eligible hospitals, and critical access hospitals (CAHs), and Medicare Advantage Organizations to promote the adoption and meaningful use of interoperable health information technology (HIT) and qualified electronic health records (EHRs). These incentive payments are part of a broader effort under the HITECH Act to accelerate the adoption of HIT and utilization of qualified EHRs.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4,847 active registrations (September, 2015)

- Name of program and sponsor Hospital Accreditation Program; The Joint Commission
- Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
- Geographic area and number and percentage of accountable entities and patients included Nationwide; 3300 Joint Commission-accredited hospitals (2014)

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Not applicable.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not applicable.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

While the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are currently available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs, one of which is this measure. Data will be accepted through February 28th, 2017.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Not applicable.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0068 : Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet
0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy
0435 : STK 02: Discharged on Antithrombotic Therapy

0438 : STK 05: Antithrombotic Therapy By End of Hospital Day Two

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy – American Academy of Neurology

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Measures 0438 Antithrombotic Therapy By End of Hospital Day 2 is the fifth (STK-5) measure in The Joint Commission stroke core measure set and also targets the ischemic stroke population; however, the timeframe for antithrombotic administration is different in this measure than STK-2. STK-5 focuses on the early management of stroke care and antithrombotic therapy administered within the first 48 hours of acute ischemic stroke onset rather than discharge. All common data elements for these measures are completely harmonized. NQF#0435: STK 02: Discharged on Antithrombotic Therapy: The measures are completely harmonized to the extent possible, given the fact that the data source for #0435 is the paper medical record, and the data source for #2832 is the electronic health record. Measure 0068 is a physician performance measure and could extend to the outpatient setting. Measure 0068 encompasses a different target population, specifically patients with ischemic vascular disease who were discharged alive for acute myocardial infarction (AMI), coronary artery bypass graft (CABG) or percutaneous coronary interventions (PCI). As previously noted, both of these measures evaluate physician practice as opposed to hospital processes.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not applicable.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2833
Measure Title: STK-03: Anticoagulation Therapy for Atrial Fibrillation/Flutter
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of ischemic stroke patients with atrial fibrillation/flutter who are prescribed anticoagulation therapy at hospital discharge.

This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-4: Thrombolytic Therapy, STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-6 Discharged on Statin Medication, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs. STK-3, Anticoagulation Therapy for Atrial Fibrillation/Flutter, is one of six of the measures in this set that have been reengineered as eCQMs and are included in the EHR Incentive Program and Hospital Inpatient Quality Reporting Program.
Developer Rationale: Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Stroke risk is significantly increased for patients with a history of or current finding of atrial fibrillation/flutter. Antiplatelet medications alone do not significantly reduce stroke risk for this patient population. Anticoagulation through the administration of warfarin, low molecular-weight heparins, or newer anticoagulant medications approved for stroke prevention is the recommended therapy.

Multiple clinical trials have demonstrated the superior therapeutic effect of warfarin compared with placebo in the prevention of thromboembolic events among patients with nonvalvular atrial fibrillation. Based on data pooled from 5 primary prevention trials of warfarin versus control, it is possible to reduce the annual stroke rate from 4.5% for the control patients to 1.4% in patients treated with dose-adjusted warfarin. In other words, 31 ischemic strokes can be prevented each year for every 1,000 patients treated (Sacco RL, et al., 2006).

Healthcare organizations that track anticoagulation therapy for internal quality improvement purposes have seen a significant increase in the measure rate over time. The chart-abstracted measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

Numerator Statement: Patients prescribed anticoagulation therapy at hospital discharge.
Denominator Statement: Patients with a principal diagnosis of ischemic stroke, history of atrial ablation, and current or history of atrial fibrillation/flutter.
Denominator Exclusions: Denominator Exclusions:

- Patients with comfort measures documented.
- Patients admitted for elective carotid intervention. This exclusion is implicitly modeled by only including non-elective hospitalizations.
- Patients discharged to another hospital.
- Patients who left against medical advice.
- Patients who expired.

- Patients discharged to home for hospice care.
- Patients discharged to a health care facility for hospice care.

Denominator Exceptions:
- Patients with a documented reason for not prescribing anticoagulation therapy at discharge.

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Pharmacy
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary:

Evidence for this new eMeasure is the same as the existing measure #0436 STK-03: Anticoagulation Therapy for Atrial Fibrillation/Flutter:

- The body of evidence consistently supports that the administration of anticoagulation therapy, unless there are contraindications, is an established and effective strategy in preventing recurrent stroke in high stroke risk atrial fibrillation patients with TIA or prior stroke.
- Class I, Level of Evidence A – 2010 AHA/ASA Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack, page 243
  - For patients with ischemic stroke or TIA with paroxysmal (intermittent) or permanent AF, anticoagulation with a vitamin K antagonist (target INR, 2.5; range 2.0 to 3.0) is recommended.
  - The Committee noted that the medical evidence to support anticoagulation therapy is not controversial;
  - however, there are some questions around the evidence for the timing of anticoagulant therapy.
- The 2012 Committee that reviewed the chart-abstracted version of this measure noted that the medical evidence to support anticoagulation therapy is not controversial; however, there are some questions around the evidence for the timing of anticoagulant therapy.

Questions for the Committee:
- *The developer attests the underlying evidence for the measure #0436 (chart-abstracted version) has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the chart-abstracted measure has not changed and can be applied to the eMeasure, therefore there is no need for repeat discussion and voting on Evidence?*

Guidance from the Evidence Algorithm: Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency: high(Box 5) → Box 5a High

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**
Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer states no performance data for the electronic version of this measure are yet available.
- This eCQM is currently included in the Hospital Inpatient Quality Reporting Program (IQR) and EHR Incentive program.
    - In CY2016, CMS is requiring organizations participating in HIQR to electronically submit one quarter of data on 4 of 28 available eCQMs. This measure is 1 of the 28 eCQMs.
- The developer also provided performance data from the chart abstracted measure which seems to be very high with little room for improvement.

Disparities:

- The developer does not provide disparities data from use of this measure.
- Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age."

Questions for the Committee:

- *Is there a gap in care that warrants a national performance eMeasure?*
- *How can this measure be used to better understand disparities in care and outcomes for certain groups?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

1a. Evidence to Support Measure Focus

Comments: **Process measure. relates indirectly to the outcome of reduced stroke and/or death (desired outcome). High quality data supports a link between institution of anticoagulation therapy for patients with atrial fibrillation and reduced stroke risks.

New data pertains more to adherence to anticoagulants, not institution (would not be captured with this measure) eg Yao, Xiaoxi, et al. "Effect of adherence to oral anticoagulants on risk of stroke and major bleeding among patients with atrial fibrillation." Journal of the American Heart Association 5.2 (2016): e003074.

**Evidence supportive

**No change of evidence - no vote needed. There is room for much more evidence as it relates to disparities.

1b. Performance Gap

Comments: **Performance data included in measure worksheet showing limited gap in overall performance on this measure. note is made of potential racial/ethnic disparities but limited data available.

**No performance data for emeasure

**No data for e version. There is encouraging reason to believe that the adoption of this measure might shed light on performance gaps in populations with disparities.

1c. High Priority (previously referred to as High Impact)

Comments: **NA

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic Clinical Data, Electronic Health Record (EHR), Imaging/Diagnostic Study, Laboratory Data, and Pharmacy Data. This is an eMeasure.

Specifications:

- HQMF specifications for eMeasure are included in the document set on SharePoint. See eMeasure Technical Review below.
- The developer states that changes have been made to the eCQM specifications to reflect revisions to the chart abstracted measure from which this measure is derived.

Questions for the Committee :

- o *Are all the data elements clearly defined? Are all appropriate codes included?*
- o *Is the logic or calculation algorithm clear?*
- o *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor review:

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously Demonstrated through Bonnie testing
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible. This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment.

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☒ **No**

Method(s) of reliability testing:

- Dataset used for testing included 22 synthetic records created in the Bonnie testing system. Measure not tested in EHR.
- Data element validity testing was performed and will count for data element reliability as well – see validity testing section.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results from the Bonnie tool demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm: Precise specifications (Box 1) →empiric reliability testing as specified (Box 2) → empiric validity testing of patient-level data (Box 3) → YES: Use rating from validity testing of patient-level data elements (Box 10)→ Assessed the measure logic only (Box 11) →Low certainty or confidence data used in measure are valid (Box 12b)→Low (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Preliminary rating for reliability: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**
☒ **Empirical validity testing of the measure score**

Validity testing method:

- The Bonnie testing tool and environment with 22 synthetic patient records were used to test the measure logic and value sets. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure.
- The numerator, denominator, exclusions/exceptions , and logic statement were tested to confirm actual results met expectations.
- The developer also conducted a review of the measure specifications to confirm the measure logic was properly expressed with the current version of the QDM and that the logic matches the clinical intent of the measure.

Validity testing results:

- The testing results from the Bonnie tool reached 100% coverage and confirmed there was a test case for each pathway of logic (negative and positive test cases).
- The measure also had a 100% passing rate which confirmed that all the test cases performed as expected.
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer submitted additional information on exclusions and meaningful differences.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Is it reasonable to assume that the impact of the exclusions will be similar for the eMeasure and the chart-abstracted measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer submitted additional information on exclusions and meaningful differences.

Question for the Committee:

- *For meaningful differences in the chart-abstracted measure the developer refers to the performance scores over time. Does the SC think the eMeasure will demonstrate similar results to the chart-abstracted measure which show little room for improvement?*

2b6. Comparability of data sources/methods: Not Applicable

2b7. Missing Data

- The developer describes the handling of missing data but does not provide information on the frequency of missing data or potential impact on measure results due to limited testing abilities using the Bonnie tool.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → threats to validity empirically assessed to some extent (Box 2) →empirical testing of data elements using BONNIE tool (Box 3,10) → Moderate

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **Low inconsistency. Target population would find reducing stroke risk and or death meaningful as long as other risks do not outweigh this (for example, major bleeding episodes)

**No concerns

**This is the e version of the existing chart measure. All appropriate data appear to be there.

2a2. Reliability Testing

Comments: **Yes. large scale validity testing 1318 hospitals and 2206379 patients. Strong Pearson correlations with 6 other measures of stroke performance.

**small sample tested; too preliminary to generalize

**22 synthetic records submitted to Bonnie test, Results are positive. Sample size is small. Implementation should allow for more detailed review - especially with regard to varied populations.

2b2. Validity Testing

Comments: **The data does not pose a threat to validity for this measure. Patients discharged to hospice or having comfort measures are usually not candidates for anticoagulation due to limited nature of active medical treatment in such population. Possible risk would be patients with dementia referred to hospice for services but for whom survival duration may be indefinite, and for home potential benefits of anticoagulation exist.
**numerous threats to validity
**Exclusions appear appropriate. Missing data is uncertain. Since this is a part of meaningful use it is likely that missing data will be minimal and thus no threats to validity.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Reliability tested in 77 hospital and 739 patients. appears sufficient.
Demonstrated inter rater reliability of 98.1% overall.
**Low reliability testing
**22 synthetic records submitted to Bonnie test. Results are good. There is a minor concern that these results are a rather small sample. Implementation should allow for much more detailed review of reliability.

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is EHR; all data elements are in defined fields in EHR.
- The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible.
- This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment.
- Per developer, they were unable to directly assess data accuracy of the data elements in an EHR system. Workflow, or the degree to which the data elements are captured during the course of care was also not assessed.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Concern is that chart review extraction is time consuming and costly for health systems or health practitioners. This kind of data is not easily obtained from EMR unless specific templates allowing data extraction are used.
**Acceptable
**Data should all be in the EHR. Accuracy and completeness of data entry is uncertain. As with all measures there is some question

as to whether the time involved in data entry and extraction for review will add value beyond what would happen without this measure.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- CMS Hospital Inpatient Quality Reporting Program (IQR)/EHR Incentive Program: The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. The EHR Incentive Program provides incentive payments to promote the adoption and meaningful use of interoperable HIT and qualified EHRs.
- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results:

- The developer states that while the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are currently available.
- In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs, one of which is this measure. Data will be accepted through February 28th, 2017.

Unexpected findings (positive or negative) during implementation:

- Developer did not identify unexpected findings during implementation.

Potential harms:

- Developer did not identify any unintended consequences related to this measure.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Noted reported in a number of public registries, websites.
since performance gap overall is low, utility for this measure would be highest in low performing systems or in populations where gap remains (example given is non-hispanic Black women.

**No performance data provided so use/usability unknown

**Part of the meaningful use accountability program. Data available and usable. Benefits to treatment are without question. Benefits of the implementation of the measure are uncertain - cost/benefit.

Criterion 5: Related and Competing Measures

Related or competing measures:

- 0436 : STK-03: Anticoagulation Therapy for Atrial Fibrillation/Flutter
- 1525 : Atrial Fibrillation and Atrial Flutter: Chronic Anticoagulation Therapy

- 0084 : Heart Failure (HF) : Warfarin Therapy Patients with Atrial Fibrillation – AMA-PCPI – no longer NQF endorsed
- 0241 : Stroke and Stroke Rehabilitation: Anticoagulant Therapy Prescribed for Atrial Fibrillation (AF) at Discharge – AAN – no longer endorsed
- 0624 : Atrial Fibrillation - Anticoagulation Therapy – ActiveHealth Management – not NQF endorsed

Harmonization:

- The target population for measure #1525 differs from measure 0436 Anticoagulation Therapy for Atrial Fibrillation/Flutter in that it includes in the denominator population all patients age 18 years and older with a diagnosis of nonvalvular atrial fibrillation or atrial flutter whose assessment of the specified thromboembolic risk factors indicate one or more high-risk factors or more than one moderate risk factor, as determined by CHADS2 risk stratification. It is not specified for ischemic stroke patients with atrial fibrillation/flutter only.
- #0436 and #2833 are completely harmonized to the extent possible, given the fact that the data source for #0436 is the paper medical record, and the data source for #2833 is the electronic health record.

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0436 NQF Project: Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The focus of the measure is to prevent a secondary stroke attributable to a thromboembolic event resulting from atrial fibrillation following a new ischemic stroke.
Anticoagulation therapy prescribed at discharge >> secondary stroke prevention >>decreased morbidity and mortality.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
This measure is consistent with the guidelines recommended by the American Heart Association/American Stroke Association for prevention of stroke in patients with ischemic stroke or transient ischemic attack. Both persistent atrial fibrillation and paroxysmal atrial fibrillation are potent predictors of first and recurrent stroke. Data from clinical trials show that age, recent congestive heart failure, hypertension, diabetes, and prior thromboembolism have been found to identify high-risk groups for atrial thromboembolism among patients with atrial fibrillation. Left ventricular dysfunction, left atrial size, mitral annular calcification (MAC), spontaneous echo contrast, and left atrial thrombus by echocardiography have also been shown to predict increased thromboemoblic risk. Overall, patients with prior stroke carry the highest stroke risk (RR, 2.5) (Sacco RL, et al., 2006).

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: Between 1990 and 2006, a total of 24 randomized trials (RCTs) involving patients with non-valvular atrial fibrillation were published. Collectively, these RCTs represent 20,012 participants with an average follow-up of 1.6 years, and a total exposure of about 32,800 patient-years (Fuster V, et al., 2006).

A meta-analysis of six randomized trials (Hart RG, et al.1999) conducted between 1989 and 1993 showed that adjusted dose-warfarin is highly efficacious for prevention of all stroke (both ischemic and hemorrhagic), with a risk reduction of 61% (95% CI 47% to 71%) versus placebo. The duration of follow-up was generally between 1 and 2 years; the longest was, 2.2 years. Five of the six RCTs, focused on primary prevention of thromboembolism in patients with nonvavlular atrial fibrillation, (i.e., The Copenhagen AFASAK Study (AFSAK); Stroke Prevention in Atrial Fibrillation (SPAF I); The Boston Area Anticoagulation Trial for Atrial Fibrillation (BAATAF); the Canadian Atrial Fibrillation Anticoagulation (CAFA) Study; and, the Veteran Administration Stroke Prevention in Non-Rheumatic Atrial Fibrillation (VA-SPINAF). One trial, the European Atrial Fibrillation Trial (EAFT), specifically studied secondary prevention among patients who had survived nondisabling stroke or cerebral transient ischemic attack (TIA). Findings from primary prevention studies were extrapolated and applied to secondary prevention.

A Cochrane review of anticoagulants for preventing stroke in patients with nonrheumatic atrial fibrillation and a history of stroke or transient ischemic attack published in The Cochrane Library 2011, Issue 4, focused specifically on two of the above RCTs: EAFT and the VA-SPINAF. The search strategy employed the Cochrane Stroke Group Trials Register (June 9, 2003), and contacted

researchers in the field to identify any further published or non-published studies. Selection criteria involved randomized trials comparing oral anticoagulants with control (no therapy) or placebo in people with NRAF and a previous TIA or minor ischemic stroke. Control groups on aspirin did not meet search criteria. These two trials involved 485 participants. Although VA-SPINAF was a primary prevention study, a small proportion of patients had suffered a previous stroke or TIA and the results for these patients were reported separately. The review considered the following main outcomes: fatal or non-fatal (disabling or non-disabling) recurrent stroke; all major vascular events: vascular death (including fatal bleeds), recurrent stroke (both ischemic and hemorrhagic), myocardial infarction, and systemic embolism; any intracranial bleed; and, major extracranial bleed defined as severe enough to lead to hospital admission, blood transfusion, or surgery (Saxena R, Koudstaal PJ, 2011).

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting anticoagulation for stroke prevention in patients with NRAF is high. The Cochrane Review of EAFT and VA-SPINAF determined that in patients with NRAF and a recent TIA or minor ischemic stroke, oral anticoagulation therapy almost halves the odds of serious vascular event. The odds of recurrent stroke, disabling as well as non-disabling, is decreased by two-thirds. In the EAFT group of patients without contraindications to anticoagulants, the annual rate of all vascular events was 8% in patients assigned to anticoagulants (n=225) versus 17% in patients assigned to placebo (n=214). The risk of stroke was reduced from 12% to 4% per year. In other words, 90 vascular events (mainly strokes) were prevented per 1000 patients treated with anticoagulants per year. The incidences of bleeding events (major or minor, intracranial or extracranial) on anticoagulation was low (2.8% per year versus 0.7% per year in the placebo group). In the VA-SPINAF study, four patients in the placebo group (n=25) compared to two in the anticoagulant group (n=21) suffered a recurrent stroke. The number of all vascular events was 8/21 in the warfarin group versus 11/25 in the placebo group (OR 0.78, 95% CI 0.20 to 2.9). No intracranial bleeds occurred. The recurrence of stroke of 9.3% per year in the placebo group, and the odds reduction by anticoagulant therapy, were very similar to those found in the EAFT.

The results of secondary prevention studies are highly similar to those of the five primary prevention studies of NRAF patients who did not have an ischemic stroke or TIA (n=4052). The most important difference between the primary and secondary prevention studies is the much higher incidence of recurrent stroke. The secondary prevention studies observed an annual incidence of 12% in the placebo group, which is nearly three times as much as in the placebo treated groups of the primary prevention studies (4.5% per year). This makes the value of anticoagulation for secondary prevention even more impressive in absolute terms: 90 vascular events (mainly strokes) are prevented if 1000 patients are treated for one year. Compared with aspirin, oral anticoagulants significantly decrease the risk of all strokes, ischemic strokes, and cardiovascular events with only a modest increase in bleeding risk. According to Van Walraven (2002), stroke risk for patients treated with aspirin alone was 10% per year for patients with a prior history of stroke and 2.7% for those with no such history versus 4% and 1.5% respectively for patients receiving anticoagulation therapy.

The Cochrane reviewers identified several flaws in EAFT. First, anticoagulant treatment was not blinded. Second, the results of the five primary prevention trials were published while EAFT was in progress. This could have biased both EAFT Auditing Committee and the individual investigators. Hospital admission for a major extracranial bleed might have been more likely for patients on an anticoagulant. However, all members of the Audit Committee who assessed the outcomes were absolutely blinded for the assigned study treatment. Finally, the majority of recurrent vascular events in EAFT were major, which left little room for inter-observer variation.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of evidence consistently supports that the administration of anticoagulation therapy, unless there are contraindications, is an established and effective strategy in preventing recurrent stroke in high stroke risk atrial fibrillation patients with TIA or prior stroke.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
There is strong evidence that oral anticoagulants are beneficial and safe for preventing a second stroke in people with atrial fibrillation and prior stroke. The literature examined reveals that patients with NRAF and a recent TIA or minor ischemic stroke who are prescribed oral anticoagulation therapy have a 50% decrease in the odds of a second serious vascular event. The odds of recurrent stroke, disabling as well as non-disabling, is also decreased by two-thirds. Patients with atrial fibrillation have an irregular heart beat and, if not properly anticoagulated, this can cause the formation of a blood clot in the left atrium of the heart. This clot may break away and block a cerebral artery causing a stroke. Patients who have had a stroke in the presence of NRAF have a high risk of another stroke. Anticoagulant drugs, such as warfarin, make the blood thinner and prevent formation of blood clots and hence could prevent stroke. However, anticoagulant drugs may also cause bleeding in the brain and this complication could offset any benefits.

However, examination of the literature reveals that many studies do not provide information on the balance between risk and benefit of anticoagulation therapy in the early period after stroke onset in patients with atrial fibrillation. Several studies have recommended withholding anticoagulants during the first few days post stroke, especially in cases of large infarcts. In one trial involving 449 patients with ischemic stroke and atrial fibrillation, a low-molecular-weight heparin did not reduce recurrent stroke risk during the first 14 days (Berge, et al, 2000). Another large trial of similar patients (N=3169) receiving subcutaneous unfractionated heparin, the risk of recurrent ischemic stroke within the first 14 days was reduced by 50%; however, an increase in spontaneous intracerebral hemorrhage offset the benefit
(Saxena, et al., 2001). Conclusions from these studies continue to recommend oral anticoagulation treatment as secondary prevention for those patients' who have no contraindication.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: American Heart Association. During our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, and provided information that was accessible and met the requirements set out in this measure maintenance form.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.13 Grade Assigned to the Body of Evidence: Class I, Level of Evidence A

1c.14 Summary of Controversy/Contradictory Evidence: There is some uncertainty as to when anticoagulation therapy should be initiated for patients with atrial fibrillation following a stroke. As noted above, one study found an increased risk of intracerebral hemorrhage when anticoagulant therapy was initiated within the first 14 days post stroke; however, it is generally recommended that anticoagulation therapy be initiated within the first one to two weeks. Although the absolute benefits of anticoagulation are higher in secondary prevention, it remains undisputed that the risk of stroke warrants the use of warfarin in all patients with atrial fibrillation, unless contraindicated. The benefit of unfractionated and low molecular-weight heparins over warfarin, as well as target INR value, is somewhat controversial.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- Berge E, Abdelnoor M, Nakstad PH, Sandset PM. Low molecular-weight heparin versus aspirin in patients with acute ischemic stroke and atrial fibrillation: a double-blinded randomized study. HAEST Study Group. Lancet. 2000;355:1205-10.
- Fuster V, Rydén LE, Cannom DS, Crijns HJ, Curtis AB, Ellenbogen KA, Halperin JL, Le Heusey JY, Kay GN, Lowe JE, Olsson SB, Prystowsky EN, Tamargo JL, Wann S; ACC/AHA Task Force Members; ESC Committee for Practice Guidleines. Guidelines for the management of patients with atrial fibrillation-executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines; developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. Circulation. 2006;114;700-752.
- Hart RJ, Benavente O, McBride R, et al. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. Ann Intern Med. 1999;131:492-501.
- Saxena R, Koudstaal PJ for The Cochrane Collaboration. Anticoagulants for preventing stroke in patients with nonrheumatic atrial fibrillation and a history of stroke or transient ischemic attack (review). The Cochrane Library. 2011:4: 1-13.
- Saxena R, Lewis S, Berge E, Sandercock PA, Koudstaal PJ. Risk of early death and recurrent stroke effect of heparin in 3169 patients with acute ischemic stroke and atrial fibrillation in the International Stroke Trial. Stroke. 2001;32:2333-7.
- Sacco RL, Adams R, Albers G, Alberts MJ, Benavente O, Furie K, Goldstein LB, Gorelick P, Halperin J, Harbaugh R, Johnston SC, Katzan I, Kelly-Hayes M, Kenton EJ, Marks M, Schwamm LH, Tomsick T. Guidelines for Prevention of Stroke in Patients With Ischemic Stroke or Transient Ischemic Attack: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association Council on Stroke: Co-Sponsored by the Council on Cardiovascular Radiology and Intervention. Stroke. Vol. 37, 2006:577.
- Van Walraven C, Hart RG, Singer DE, Laupacis A, Connolly S, Petersen P, Koudstaal PJ, Chang Y, Hellemons B. Oral anticoagulants vs. aspirin in nonvalvular atrial fibrillation: an individual patient meta-analysis. JAMA. 2002;288:2441-8.

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)*:
2010 AHA/ASA Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack, page 243
Class I, Level of Evidence A

For patients with ischemic stroke or TIA with paroxysmal (intermittent) or permanent AF, anticoagulation with a vitamin K antagonist (target INR, 2.5; range 2.0 to 3.0) is recommended.

1c.17 Clinical Practice Guideline Citation: Furie KL, Kasner SE, Adams RJ, Albers GW, et al. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2011;42:241-43.

1c.18 National Guideline Clearinghouse or other URL: http://stroke.ahajournals.org/content/42/1/227.full.pdf

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: This guideline was approved by the American Heart Association Science Advisory and Coordinating Committee on July 13, 2010. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.23 Grade Assigned to the Recommendation: Class I, Level of Evidence A

1c.24 Rationale for Using this Guideline Over Others: The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: High 1c.26 Quality: High1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0436_Evidence_MSF5.0_Data-635827722551306673.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Stroke risk is significantly increased for patients with a history of or current finding of atrial fibrillation/flutter. Antiplatelet medications alone do not significantly reduce stroke risk for this patient population. Anticoagulation through the administration of warfarin, low molecular-weight heparins, or newer anticoagulant medications approved for stroke prevention is the recommended therapy.

Multiple clinical trials have demonstrated the superior therapeutic effect of warfarin compared with placebo in the prevention of thromboembolic events among patients with nonvalvular atrial fibrillation. Based on data pooled from 5 primary prevention trials of warfarin versus control, it is possible to reduce the annual stroke rate from 4.5% for the control patients to 1.4% in patients treated with dose-adjusted warfarin. In other words, 31 ischemic strokes can be prevented each year for every 1,000 patients treated (Sacco RL, et al., 2006).

Healthcare organizations that track anticoagulation therapy for internal quality improvement purposes have seen a significant increase in the measure rate over time. The chart-abstracted measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This measure is a legacy eCQM that is currently included in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data for the electronic version of the measure are yet available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs. This measure is one of the 28.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
In October, 2009, The Joint Commission added the chart-abstracted stroke (STK) measure set as a new core measure option to meet performance measure requirements for Joint Commission hospital accreditation purposes. Below is the specified level of analysis for STK-3 beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 325 denominator cases; 303 numerator cases; 43 hospitals; 0.93231 national aggregate rate; 0.90522 mean of hospital rates; 0.22409 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.875 25th percentile rate/lower quartile; and, 0.8 10th percentile rate.

CY 2010: 2952 denominator cases; 2785 numerator cases; 136 hospitals; 0.94343 national aggregate rate; 0.92168 mean of hospital

rates; 0.13488 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.90767 25th percentile rate/lower quartile; and, 0.66667 10th percentile rate.

CY 2011: 3566 denominator cases; 3381 numerator cases; 150 hospitals; 0.94812 national aggregate rate; 0.92886 mean of hospital rates; 0.13698 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.90625 25th percentile rate/lower quartile; and, 0.7735 10th percentile rate.

CY 2012: 3685 denominator cases; 3530 numerator cases; 149 hospitals; 0.95794 national aggregate rate; 0.94795 mean of hospital rates; 0.11113 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.95522 25th percentile rate/lower quartile; and, 0.83333 10th percentile rate.

CY 2013: 5635 denominator cases; 5429 numerator cases; 257 hospitals; 0.96344 national aggregate rate; 0.95363 mean of hospital rates; 0.1165 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.95349 25th percentile rate/lower quartile; and, 0.88235 10th percentile rate.

CY 2014: 28,027 denominator cases; 27,261 numerator cases; 1256 hospitals; 0.97267 national aggregate rate; 0.96598 mean of hospital rates; 0.08548 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.96429 25th percentile rate/lower quartile; and, 0.88889 10th percentile rate.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al., 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

In the national REasons for Geographic and Racial Differences in Stroke (REGARDS) cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance

and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years of age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age : RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer,et al., 2010).

In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).
Although blacks and African Americans have a lower incidence of atrial fibrillation than Non-Hispanic whites (Hajat, et al. 2001), blacks and African Americans are also less likely to undergo cardiac monitoring and noninvasive cerebrovascular testing (Mitchell, et al., 2000). In REGARDS, investigators found that blacks or African Americans were less likely to be aware that they had atrial fibrillation or to be treated with warfarin.

Furthermore, minorities are less likely to receive medications for secondary prevention. One report suggests that blacks or African Americans are less likely to have thorough diagnostic evaluation after first stroke and are less likely to receive guideline-concordant stroke preventive medications, such as warfarin or other anticoagulants. In another study which used the 2005 Behavioral Risk Factor Surveillance System (BRFSS) in 11, 862 stroke survivors, little difference was found among blacks or African Americans and non-Hispanic whites in terms of secondary prevention measures. The study found that secondary prevention measures were underutilized in both racial groups.

Studies have also noted a relationship between health literacy, particularly math skills and medication compliance. A study from Estrada and colleagues (2004), found that anticoagulation control was poorer for participants with lower literacy levels. The international normalized ratio (INR) was 32% higher for participants in the lowest literacy group versus the highest (P+0.009). Other studies have found no association between literacy and the proportion of time with the INR in the therapeutic range (OR 1.0, 95% CI 0.7 to 1.4); however, no genetic factors influencing response to anticoagulation were included in the analysis (Fang MC, et al., 2006).

Since the last endorsement date, Schwamm and colleagues (2010) reported that black patients had significantly lower adjusted odds compared with white patients of receiving anticoagulation for atrial fibrillation (OR, 0.84; 95% CI, 0.75 to 0.94). Findings from Qian and associates (2013) agreed that non-Hispanic black patients were less likely to receive anticoagulation for atrial fibrillation at discharge. Using patient data (n=200,900) from the American Heart Association/American Stroke Association Get With The Guidelines (GWTG)-Stroke program from April 2003 through December 2008, Qian reported the following performance measure rates for discharged on anticoagulation for atrial fibrillation: non-Hispanic White (n=170,694) 90.0%; non-Hispanic Black (n=20,514) 88.3%; Hispanic (n=6632) 89.0%; and non-Hispanic Asian American (n=3060) 90.3%. According to data from the Paul Coverdell National Acute Stroke Registry (PCNASR) (n=9358), patients who are not white are less likely to receive anticoagulation therapy for atrial fibrillation; White 96.2%; Other Race 94.0% (Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, 2014).

White women with atrial fibrillation, as well as women of other races with atrial fibrillation, are slightly less likely to receive anticoagulation therapy than men (88% versus 89.7%; adjusted OR, 0.93; 95% CI, 0.88–0.98) (Bushnell, 2014). The attributable risk of stroke from atrial fibrillation increases with age, from 1.5% for those aged 50 to 59 years to nearly 25% for those aged =80 years. Whites carry the highest prevalence of atrial fibrillation compared with blacks, Hispanics, Asians, and other ethnic groups. The overall number of men and women with atrial fibrillation is similar, but ˜60% of atrial fibrillation patients aged >75 years are women

(15.6% of men and 20.4% of women (P<0.0001).

• Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, Yan T, Escarce JJ. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-8.
• Bushnell, C, McCullough LD, Awad IA, Chireau MV, Fedder WN, Furie KL, Howard VJ, Lichtman JH, Lisabeth LD, Piña IL, Teeves MJ, Rexrode KM, Saposnik G, Singh V, Towfighi A, Vaccarino V, Walters MR. Guidelines for the Prevention of Stroke in Women A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association The American Academy of Neurology affirms the value of this guideline as an educational tool for neurologists. Stroke. 2014;45:24-25.
• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, Griffith P, Gorelick PB, Howard G, Leira EC, Morgenstern LB, Ovbiagele B, Peterson E, Rosamond W, Trimble B, Valderrama AL, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• Estrada CA, Martin-Hryniewicz M, Peek BT, Collins C, Byrd JC. Literacy and numeracy skills and anticoagulation control. Am J Med Sci. 2004;328:88-93.
• Fang MC, Machtinger EL, Wang F, Schillinger D. Health literacy and anticoagulation-related outcomes among patients taking warfarin. J Gen Intern Med. 2006; 21:841-46.
• Gillum RF, Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 2011;42(11):3294-6.
• Hajat C, Dundas R, Stewart JA< Lawrence E, Rudd AG, Howard R, Wolf CD. Cerebrovascular risk factors and stroke subtypes: differences between ethnic groups. Stroke. 2001;32:37-42.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-75.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Karve S, Balkrishnan R, Seiber E Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Levine DA, Kiefe CI, Howard G, Howard VJ, Williams OD, Allison JJ. Reduced medication access: a marker for vulnerability in US stroke survivors. Stroke. 2007;38:1557-64.
• Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Mitchell JB, Ballard DJ, Atchar DB, Whisant JP, Samsa GP. Racial variation in treatment for transient ischemic attacks: impact of participation by neurologists. Heatlh Serv Res. 2000;34:1413-1428.
• Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, Kerber KA, Skolarus LE, Meurer WJ, Burke JF, Adelman EE, Baek J, Lisabeth LD. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol. 2013;74:778–785.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Ross JS, Halm EA, Bravata DM. Use of stroke secondary prevention services: are there disparities in care? Stroke. 2009;40:1811-19.
• Schwamm LH, Syed FA, Reeves MJ, Smith EE, Saver JL, Messe S, Bhatt DL, Grau-Sepulveda MV, Peterson ED, Fonarow GC. Temporal

trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With the Guidelines-Stroke hospitals. Circ Cardiovasc Quality Outcomes. 2013;6:543-549
• Tuhrim S, Cooperman A, Rojas M, Brust JC, Koppel B, Martin K, Chassin M. The association of race and sex wih the underuse of stroke prevention measures. J Cerebrovascular Dis. 2008;17:226-234.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States is still attributable to stroke.. More women than men die of stroke each year. Women accounted for almost 60% of US stroke deaths in 2008 (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 ($P<0.05$) (US Burden of Disease Collaborators, 2013) . Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

Approximately 20% of ischemic strokes result from a cerebral embolism secondary to a cardiac arrhythmia or disorder. Atrial fibrillation (AF) is the most common sustained cardiac rhythm disturbance (CDC, 2010). Paroxysmal, persistent, and permanent atrial fibrillation are strong predictors of first and recurrent stroke, increasing ischemic stroke risk four to five-fold. It is estimated that over 2.3 million Americans have atrial fibrillation, and the incidence becomes more prevalent with age. AF accounts for ˜ 1.5% of stroke in individuals 50 to 59 years of age to nearly 25% in those aged > 80 years (Bushnell C, et al., 2014).

Patients who have suffered an ischemic stroke who have a high-risk source of cardiogenic embolism should generally be treated with anticoagulant drugs to prevent reoccurrence. For most patients with ischemic stroke and atrial fibrillation, it is reasonable to initiate anticoagulation therapy within 14 days of stroke onset (Kernan WN, eta l, 2014). Warfarin, dabigatran, and apixaban are all indicate for the prevention of recurrent stroke in patients with nonvalvular atrial fibrillation, whether paroxysmal or permanent. Rivaroxaban is a reasonable alternative (Kernan WN, 2014). Ischemic stroke rates per 1000 patient-years declined in AF patients taking anticoagulants, from 46.7% in 1992 to 19.1% in 2002 (AHA 2012). According to the Framingham Study (1996), AF is also an independent risk factor for ischemic stroke severity, recurrence, and mortality (Lin HJ, et al., 1996). In a study from Penado and associates (2003), people who had AF and were not treated with anticoagulants had a 2.1-fold increase in risk for recurrent stroke and a 2.4-fold increase in risk for recurrent severe stroke.

In addition to the costs attributed to stroke, the treatment of atrial fibrillation alone represents a significant health care burden. The estimated cost of treatment of atrial fibrillation in 2005 was $6.65 billion per year, including the costs of hospitalization, inpatient and outpatient physician care, and medications (Roger VL, et al.,2012).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
• Bushnell C, McCullough LD, Awad IA, Chireau MV, Fedder WN, Furie KL, Howard VJ, Lichtman JH, Lisabeth LD, Piña IL, Teeves MJ, Rexrode KM, Saposnik G, Singh V, Towfighi A, Vaccarino V, Walters MR, Guidelines for the Prevention of Stroke in Women A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association The American Academy of

Neurology affirms the value of this guideline as an educational tool for neurologists. Stroke. 2014;45:24-25.
• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Centers for Disease Control and Prevention, Division for Heart Disease and Stroke Prevention, National Center for Chronic Disease Prevention and Health Promotion. Fact sheets and at-a-glance reports: atrial fibrillation fact sheet. 2010.
• George M., Xin Tong, McGruder H., Yoon P., Rosamond W., Winquist A., Hinchey J., Wall H., Pandey D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, Fang MC, Fisher M, Furie KL, Heck DV, Johnston SC, Kasner SE, Kittner SJ, mitchell PH, Rich MW, Richardson D, Schwamm LH, Wilson JA. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2014;45:241-2160-2236.
• Lin HJ, Wolf PA, Kelly-Hayes M, Beiser AS, Kase CS, Benjamin EJ, D'Agostino RB. Stroke severity in atrial fibrillation: The Framingham Study. Stroke. 1996;27:1760-64.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Penado S, Cano M, Acha O, Hernández JL,Riancho JA. Atrial fibrillation as a risk factor for stroke recurrence. Am J Med. 2003;114:206-210.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e-78-82, e-119-127.

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input was obtained.)*

Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/ecqm_library.html

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is an eMeasure **Attachment:** STK3_MAT.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** STK3Stroke_v5_Wed_Apr_01_12.13.27_CDT_2015.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Changes have been made to the eCQM specifications in order to reflect revisions to the chart abstracted measure from which this measure is derived.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients prescribed anticoagulation therapy at hospital discharge.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Anticoagulation Therapy Prescribed at Discharge
- Anticoagulant Therapy is represented with the QDM datatype and value set of Medication, Discharge: Anticoagulant Therapy (OID: 2.16.840.1.113883.3.117.1.7.1.200)

Non-Elective Inpatient Encounter
- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Patients with a principal diagnosis of ischemic stroke, history of atrial ablation, and current or history of atrial fibrillation/flutter.

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Principal Diagnosis of Ischemic Stroke

- Ischemic Stroke is represented with the QDM datatype and value set of Diagnosis, Active: Ischemic Stroke (OID: 2.16.840.1.113883.3.117.1.7.1.247)
- Ordinality: Principal (OID: 2.16.840.1.113883.3.117.1.7.1.14)

Non-Elective Inpatient Encounter

- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

History of Atrial Ablation

- Atrial Ablation is represented with the QDM datatype and value set of Procedure, Performed: Atrial Ablation (OID: 2.16.840.1.113883.3.117.1.7.1.203)

Current or Historical Diagnosis of Atrial Fibrillation/Flutter

- Current Diagnosis of Atrial Fibrillation/Flutter is represented with the QDM datatype and value set of Diagnosis, Active: Atrial Fibrillation/Flutter (OID: 2.16.840.1.113883.3.117.1.7.1.202)
- Historical Diagnosis of Atrial Fibrillation/Flutter is represented with the QDM datatype and value set of Diagnosis, Inactive: Atrial Fibrillation/Flutter (OID: 2.16.840.1.113883.3.117.1.7.1.202)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Denominator Exclusions:

- Patients with comfort measures documented.
- Patients admitted for elective carotid intervention. This exclusion is implicitly modeled by only including non-elective hospitalizations.
- Patients discharged to another hospital.
- Patients who left against medical advice.
- Patients who expired.
- Patients discharged to home for hospice care.
- Patients discharged to a health care facility for hospice care.

Denominator Exceptions:

- Patients with a documented reason for not prescribing anticoagulation therapy at discharge.

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Denominator Exclusions Data Elements:
Non-Elective Inpatient Encounter

- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

Discharge Status (modeled as Attributes of the above Non-Elective Inpatient Encounter)
- Discharge status: Left Against Medical Advice (OID: 2.16.840.1.113883.3.117.1.7.1.308)
- Discharge status: Patient Expired (OID: 2.16.840.1.113883.3.117.1.7.1.309)
- Discharge status: Discharge To Acute Care Facility (OID: 2.16.840.1.113883.3.117.1.7.1.87)
- Discharge status: Discharged to Home for Hospice Care (OID: 2.16.840.1.113883.3.117.1.7.1.209)
- Discharge status: Discharged to Health Care Facility for Hospice Care (OID: 2.16.840.1.113883.3.117.1.7.1.207)

Comfort Measures
- Comfort Measures are represented with the QDM datatypes and value set of:
- Intervention, Order: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)
- Intervention, Performed: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)

Emergency Department Visit
- Emergency Department Visit is represented with the QDM datatype and value set of Encounter, Performed: Emergency Department Visit (OID: 2.16.840.1.113883.3.117.1.7.1.292)

Denominator Exceptions Data Elements:
Reasons for Not Prescribing Anticoagulation Therapy
- Anticoagulant Ingredient Specific Medication is represented with the QDM datatype and value set of Medication, Discharge: Anticoagulant ingredient specific (OID: 2.16.840.1.113762.1.4.1021.9)
- Medical Reason is represented with the QDM datatype and value set of Medication, Discharge not done: Medical Reason (OID: 2.16.840.1.113883.3.117.1.7.1.473)
- Patient Refusal is represented with the QDM datatype and value set of Medication, Discharge not done: Patient Refusal (OID: 2.16.840.1.113883.3.117.1.7.1.93)

Non-Elective Inpatient Encounter
- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable, the measure is not stratified.

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable.

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. Type of score:
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
See attached HQMF file.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
eMeasures are calculated using only the structured data collected in certified EHR technology (CEHRT). Data not present in the structured field from which the measure draws will not be included in the measure calculation.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Pharmacy

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Hospitals report EHR data using Certified Electronic Health Record Technology (CEHRT), and by submitting Quality Reporting Document Architecture Category 1 (QRDA-1).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF0436_CMS71v4_STK3_Bonnie_Testing.xlsx,STK3_eCQM_testing_attachment-635907890064535446.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2833
Measure Title: STK03: Anticoagulation Therapy for Atrial Fibrillation/Flutter
Date of Submission: 1/14/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	X Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the sub criteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure*. *Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**

☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒other: Bonnie Test Cases	☒ other: Bonnie Test Cases (see question 1.2)

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

This measure is a legacy eCQM (an NQF endorsed measure that has been re-specified into an eMeasure and is currently used in a federal quality program/programs). In accordance with the modified NQF requirements for testing of eMeasures, the Bonnie testing tool was used to simulate a testing environment where measure specifications and HQMF output are tested against synthetic test data. Measure developers rely on the results in Bonnie to confirm whether the measure logic is performing as expected. Reference the eCQI Resource Center website (https://ecqi.healthit.gov/ecqm-tools/tool-library/bonnie) or the Bonnie testing tool website (https://bonnie.healthit.gov/) for more information about Bonnie functionality and its role in measure development. Please also reference the Bonnie testing worksheet attachment for detailed Bonnie test cases and testing results for this measure.

While Bonnie does consume HQMF specifications, we are not able to test the measure with data from HQMF implemented in EHRs.

1.3. What are the dates of the data used in testing? The measure specifications were tested in the Bonnie testing environment which mimics the year 2012.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☒ other: Bonnie Test Cases

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Not applicable.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

22 unique synthetic patient records were created in the BONNIE testing system for this measure. Cases were used to test the validity of each data element and timing relationship in the measure. Patient characteristics such as age, diagnosis, and medications were pre-determined to provide a variety of scenarios that adequately test patients passing

each data element and failing each data element. Data included in cases and tested for this measure included diagnoses, medications, discharge statuses, patient orders, procedures, age, length of stay, and ED and inpatient encounters.

All 22 cases passed or failed as expected based on the data included in the case, confirming the measure logic is accurate and valid.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Not applicable. Bonnie test patients were developed for use across the different aspects of testing.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Not applicable.

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*May be one or both levels*)

X **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

The chart-abstracted version of this measure has been in national use since the 4th quarter of 2009. At the time the chart-abstracted measure was originally tested, extensive tests of measure reliability were conducted. At present, no performance data for the electronic version of the measure are currently available. Below is the reliability testing summary for NQF #04376 STK03: Anticoagulation Therapy for Atrial Fibrillation/Flutter, from which this measure is derived.

At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with monthly data transmission to The Joint Commission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of approximately 900 patient records identified at 30 pilot sites over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rates of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify potential measure enhancements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program

which requires data collection and reporting on standardized national performance measures.) The following is a list of the major tests conducted on the submitted data for this measure:

- Transmission of complete data
- Usage of data received (to understand if the healthcare organization provides the relevant service to treat the relevant population)
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity
- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors is as follows:

- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e*.g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data element agreement rates reported to The Joint Commission for the chart-abstracted version of this measure for the time period of one year (4Q2010 to 3Q2011) have shown an overall agreement rate of 98.1%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-3.

Data Elements	total 'n' numerator	total 'n'denominator	rate
Anticoagulation Therapy Prescribed at Discharge	127	150	84.7%
Atrial Fibrillation/Flutter	446	467	95.5%
Clinical Trial	712	714	99.7%
Comfort Measures Only	709	714	99.3%
Elective Carotid Intervention	711	714	99.6%
Reason for Not Prescribing Anticoagulation Therapy at Discharge	38	46	82.6%

These agreement rates are considered to be well within acceptable levels.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Agreement rates for individual data elements tested for the chart-abstracted version of this measure were within acceptable levels. Once data are available for analysis, it is expected that reliability tests of the eCQM version of this measure will yield similar results.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The Bonnie testing tool and environment were used to test the measure logic and value sets. Each data element and logic statement was tested to confirm actual results met expectations. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure. An example of negative testing would be to include test cases with pediatric ages to ensure that pediatric patients are not included in the measure.

Denominator test cases positively test to ensure patients with principal diagnosis of ischemic stroke and either a history of atrial ablation or diagnosis of atrial fibrillation/flutter appropriately fall into the measure. We have created negative test cases, testing to ensure patients without a principal diagnosis of ischemic stroke and either a history of atrial ablation or diagnosis of atrial fibrillation/flutter do not actually make it into the denominator.

Numerator test cases positively test to ensure patients receiving anticoagulant therapy at discharge pass the measure. Negative test cases ensure that a patient who does not receive anticoagulant therapy in the specified time frame do not make pass the measure.

Denominator exclusion and exception test cases for this measure ensure that patients are properly removed from the denominator if they have comfort measures, specific discharge statuses other than discharged home, and medical reasons or patient refusal for not getting anticoagulant therapy. Negative test cases are also run, an example of this would be comfort measures that are ordered but not in the specified time frame. Testing confirmed patients meeting the exclusion and exception criteria are removed from the measure appropriately.

A review of the measure specifications was also conducted to confirm the logic was properly expressed within the current version of the QDM and confirmed the logic matches the clinical intent of the measure, as stated in the measure header. This is done through use of a logic checklist to facilitate review of the measure logic, according to several checks, a few examples of which are included below:

- Is the intent of the measure described in the measure description articulated/ captured in the measure logic?
- Do the logic elements map to definitions in the measure narrative, data dictionary or supporting reference documentation?
- Do the populations in the narrative align with the populations defined in the logic?
- Are the mathematic inequalities reflective of the measure intent and represent the intended populations (for example: when intended the inequality represents less than rather than less and or equal to)?

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Bonnie results provide coverage and passing rates. This measure reached 100% coverage, confirming there is a test case for each pathway of logic. This measure also has a 100% passing rate, confirming all test cases performed as expected.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

This measure performed as expected in Bonnie.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Exclusions in the eCQM align with the chart-based version of the measure, and are clinically necessary for the interpretation of the measure. As noted previously, the chart-abstracted version of this measure has been in national use since the 4th quarter of 2009, and no data are available for the eCQM version of the measure. The below analysis addresses exclusions testing performed for the chart-abstracted version of the measure from which this measure is derived.

The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patient who have a Length of Stay greater than 120 days
- Patients with Comfort Measures
- Patient enrolled in a Clinical Trial
- Patients admitted for Elective Carotid Intervention
- Patients with a documented Reason For Not Prescribing Anticoagulation Therapy at Discharge
- Patients discharged to another hospital
- Patients who left against medical advice
- Patients who expired
- Patients discharged to home for hospice care
- Patients discharged to a health care facility for hospice care

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are given below.

Exclusion: Comfort Measures Only
Overall Occurrence n = 163,225
Overall Occurrence Percentage 10.4%
Minimum 0.30%
10th Percentile: 5.26%
Median: 10%
90th Percentile: 15.8%
Maximum: 35.7%

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%

10th Percentile: 0.24%
Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for Elective Carotid Intervention
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.5%
90th Percentile: 25.3%
Maximum: 95.2%

Exclusion: Patients with a documented Reason For Not Prescribing Anticoagulation Therapy at Discharge
Overall Occurrence n = 25,249
Overall Occurrence Percentage 8.01%
Minimum: 0.25%
10th Percentile: 1.96%
Median: 5.22%
90th Percentile: 16%
Maximum 84.46%

Exclusion: Discharge Disposition - Patients discharged to another hospital
Overall Occurrence n = 449,924
Overall Occurrence Percentage 35.7%
Minimum: 0.787%
10th Percentile: 25%
Median: 35.4%
90th Percentile: 46%
Maximum: 76.2%

Exclusion: Discharge Disposition - Patients who left against medical advice
Overall Occurrence n = 8,396
Overall Occurrence Percentage 0.67%
Minimum: 0.067%
10th Percentile: 0.34%
Median: 0.86%
90th Percentile: 2.4%
Maximum: 9.67%

Exclusion: Discharge Disposition - Patients who expired
Overall Occurrence n = 76,168
Overall Occurrence Percentage 6.04%
Minimum: 0.398%
10th Percentile: 1.9%
Median: 5.08%
90th Percentile: 10.2%
Maximum: 20.1%

Exclusion: Discharge Disposition- Patients discharged to home for hospice care
Overall Occurrence n = 658,264

Overall Occurrence Percentage 52.2%
Minimum: 6.25%
10th Percentile: 39%

Median: 51.9%
90th Percentile: 64%
Maximum 94.3%

Exclusion: Discharge Disposition - Patients discharged to a health care facility for hospice care
Overall Occurrence n = 37,804
Overall Occurrence Percentage 3%
Minimum: 0.169%
10th Percentile: 0.87 %
Median: 3.01%
90th Percentile: 6.4%
Maximum: 23%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Analysis of these data for the chart-abstracted progenitor of this measure indicated that all exclusions were appropriate. It is believed that results for exclusions in the eCQM will be similar when sufficient data have been received to perform such an analysis.

_____

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

- ☒ **No risk adjustment or stratification**
- ☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
- ☐ **Stratification by** Click here to enter number of categories **risk categories**
- ☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of an HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

A similar methodology will be used for the eCQMs.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

NQF#0436: STK-3 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10th Percentile

3Q2011	131	0.94408	0.16256	1	1	1	1	0.83333
2Q2011	131	0.94408	0.16256	1	1	1	1	0.83333
1Q2011	147	0.92518	0.17565	1	1	1	0.94737	0.66667
4Q2010	126	0.92909	0.19535	1	1	1	1	.8
3Q2010	119	0.94605	0.15078	1	1	1	1	0.75
2Q2010	113	0.93581	0.16405	1	1	1	0.93333	0.81818
1Q2010	90	0.91729	0.15775	1	1	1	0.88889	0.66667
4Q2009	43	0.90522	0.22409	1	1	1	0.875	0.8

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

It should be noted that since data collection on this measure is completely voluntary for both The Joint Commission, the hospitals reporting on this measure are self-selected, and therefore, presumably have a particular interest and concern for improving stroke care. For this reason, the measure results demonstrated by this group most likely significantly overstates the rate for the population of all health care organizations.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Multiple data sources are not used for this measure.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Data not present in the structured field from which the measure draws will not be included in the measure calculation. In the Bonnie testing environment, missing data are tested as an expected "Fail" for that data element, and actual performance (whether or not the case fails) is compared to expectations to ensure missing data impacts data element and overall measure calculation as expected. This is the extent of missing data analysis that can be performed in Bonnie.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

See response for question 2b7.1 above.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

The measure performed as expected in the Bonnie testing environment.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment **Attachment:** STK3_eCQM_NQF_Measure_Feasibility_Assessment_Report.docx

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Not applicable.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

Value sets are housed in the Value Set Authority Center (VSAC), which is provided by the National Library of Medicine (NLM), in coordination with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

Viewing or downloading value sets requires a free Unified Medical Language System® (UMLS) Metathesaurus License, due to usage restrictions on some of the codes included in the value sets. Individuals interested in accessing value set content can request a UMLS license at (https://uts.nlm.nih.gov/license.html)

There are no other fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program http://www.jointcommission.org/ EHR Incentive Program https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprogram

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

- Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
- Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 hospitals (2015)

- Name of program and sponsor: EHR Incentive Program; Centers for Medicare & Medicaid Services
- Purpose: The American Recovery and Reinvestment Act of 2009 (ARRA) (Pub.L. 111–5) was enacted on February 17, 2009. Title IV of Division B of ARRA amends Titles XVIII and XIX of the Social Security Act (the Act) by establishing incentive payments to eligible professionals (EPs), eligible hospitals, and critical access hospitals (CAHs), and Medicare Advantage Organizations to promote the adoption and meaningful use of interoperable health information technology (HIT) and qualified electronic health records (EHRs). These incentive payments are part of a broader effort under the HITECH Act to accelerate the adoption of HIT and utilization of qualified EHRs.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4,847 active registrations (September, 2015)

- Name of program and sponsor Hospital Accreditation Program; The Joint Commission
- Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
- Geographic area and number and percentage of accountable entities and patients included Nationwide; 3300 Joint Commission-accredited hospitals (2014)

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not applicable.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

While the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are currently available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs, one of which is this measure. Data will be accepted through February 28th, 2017.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Not applicable.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

0084 : Heart Failure (HF) : Warfarin Therapy Patients with Atrial Fibrillation
0241 : Stroke and Stroke Rehabilitation: Anticoagulant Therapy Prescribed for Atrial Fibrillation (AF) at Discharge
0436 : STK-03: Anticoagulation Therapy for Atrial Fibrillation/Flutter
0624 : Atrial Fibrillation - Anticoagulation Therapy
1525 : Atrial Fibrillation and Atrial Flutter: Chronic Anticoagulation Therapy

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

0624 : Atrial Fibrillation - Anticoagulation Therapy; Active Health Management – no longer NQF-endorsed.
0084 : Heart Failure (HF) : Warfarin Therapy Patients with Atrial Fibrillation –status unspecified; AMAPCPI
0241 : Stroke and Stroke Rehabilitation: Anticoagulant Therapy Prescribed for Atrial Fibrillation (AF) at Discharge- status unspecified; AAN

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?

No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

Measure 1525 from the American College of Cardiology is a physician performance measures identified through CPT codes and could extend to the outpatient setting. The measure evaluates physician practice as opposed to hospital processes. The target population for measure 1525 differs from measure 0436 Anticoagulation Therapy for Atrial Fibrillation/Flutter in that it includes in the denominator population all patients age 18 years and older with a diagnosis of nonvalvular atrial fibrillation or atrial flutter whose assessment of the specified thromboembolic risk factors indicate one or more high-risk factors or more than one moderate risk factor, as determined by CHADS2 risk stratification. It is not specified for ischemic stroke patients with atrial fibrillation/flutter only. NQF#0436: STK-03: Anticoagulation Therapy for Atrial Fibrillation/Flutter: The measures are completely harmonized to the extent possible, given the fact that the data source for #0436 is the paper medical record, and the data source for #2833 is the electronic health record.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Not applicable.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Ann, Watt, awatt@jointcommission.org, 630-792-5944-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Lisa, Anderson, landerson2@jointcommission.org, 630-792-5008-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel (TAP) is to provide advisory oversight in literature review, measure content and maintenance of the specifications.
Members are:

Harold P. Adams, Jr., MD
University of Iowa Health Care
Iowa City, IA

Mark J. Alberts, MD
University of Texas Southwestern
Dallas, TX

Anne W. Alexandrov, RN
Health Outcomes Institute
Fountain Hills, AZ

Nadine Allyn, RD
American Heart Association
Dallas, Texas

Mary G. George, MD
Centers for Disease Control and Prevention
Atlanta, GA

Martin Gizzi, MD
Legacy Health System
Portland, OR

Marianna Gorbaty
Mathematica Policy Research, Inc.
Washington, DC

Judith Hinchey, MD
Tuft's Medical School
Steward St. Elizabeth's Medical Center
Boston, MA

Edward C. Jauch, MD
Medical University of South Carolina
Charleston, SC 2946

Irene Katzan, MD
Cleveland Clinic
Cleveland, OH

Kimberly E. Levasseur-Franklin, PharmD
Northwestern Memorial Hospital
Chicago, IL

Kathy Morrison, RN
Penn State Hershey Medical Center
Hershey, PA

Marilyn M. Rymer, MD
University of Kansas
Kansas City, MO

Jeffrey Saver, MD
UCLA Medical Center
Los Angeles, CA

Lee H. Schwamm, MD
Harvard Medical School
Massachusetts General Hospital
Boston, MA

Deborah, Summers, RN
St. Luke's Brain and Stroke Institute
Kansas City, MO

Osama O. Zaidat, MD
St.Vincent Mercy Hospital
Toledo, OH

Richard D. Zorowitz, MD
Medstar National Rehabilitation Network
Washington, DC

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2012

Ad.3 **Month and Year of most recent revision:** 04, 2015

Ad.4 **What is your frequency for review/update of this measure?** Annual

Ad.5 **When is the next scheduled review/update for this measure?** 04, 2016

Ad.7 **Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. The measures and specifications are provided without warranty.

Ad.8 **Additional Information/Comments:** Not applicable.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2834
Measure Title: STK 04: Thrombolytic Therapy
Measure Steward: The Joint Commission

Brief Description of Measure: This measure captures the proportion of acute ischemic stroke patients who arrive at this hospital within 2 hours of time last known well for whom IV t-PA was initiated at this hospital within 3 hours of time last known well. This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-6 Discharged on Statin Medication, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs.This measure captures the proportion of acute ischemic stroke patients who arrive at this hospital within 2 hours of time last known well for whom IV t-PA was initiated at this hospital within 3 hours of time last known well.

This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-6 Discharged on Statin Medication, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs. STK-4, Thrombolytic Therapy, is one of six of the measures in this set that have been reengineered as eCQMs and are included in the EHR Incentive Program and Hospital Inpatient Quality Reporting Program.

Developer Rationale: Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Studies have demonstrated that the early administration of thrombolytic therapy within 3 hours of stroke symptom onset can significantly improve neurologic outcomes at 3 months in patients with ischemic stroke. The European Cooperative Acute Stroke Study (ECASS) III trial indicated that intravenous rtPA can be given safely to, and can improve outcomes for, carefully selected patients treated 3 to 4.5 hours after stroke; however, as the NINDS investigators concluded, the earlier that IV thrombolytic therapy is initiated, the better the patient outcome. Despite strong recommendations from the American Academy of Neurology, the American Heart Association, and the American College of Chest Surgeons, thrombolytic therapy is used in only a small proportion of ischemic stroke patients overall and in only a minority of eligible candidates.

Healthcare organizations that track IV thrombolytic administration for internal quality improvement purposes have seen a significant increase in the measure rate over time. The chart-abstracted measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

Numerator Statement: Acute ischemic stroke patients for whom IV thrombolytic therapy was initiated at this hospital within 3 hours (less than or equal to 180 minutes) of when it was witnessed or reported that the patient was last known to be without the signs and symptoms of the current stroke or at his or her baseline state of health.

Denominator Statement: Ischemic stroke patients admitted through the Emergency Department whose time of arrival is within 2 hours (less than or equal to 120 minutes) of the 1) time they were known to be at their baseline state of health; or 2) time of symptom onset if time last known at baseline state is not known.

Denominator Exclusions: Denominator Exclusions:
None.

Denominator Exceptions:
- Patients with comfort measures documented on the day of or the day after arrival

- Patients with intra-venous or intra-arterial Thrombolytic (t-PA) Therapy prior to arrival
- Patients with documentation of a National Institutes for Health Stroke Scale (NIHSS) score of zero in the emergency department
- Patients with Medical Reasons for not initiating IV thrombolytics documented by a physician/APN/PA or pharmacist on the day of or the day after arrival
- Patients with any of the following results within 180 minutes of the 1) time they were known to be at their baseline state of health; or 2) time of symptom onset:
  - Prothrombin Time > 15 seconds
  - Platelet Count <100,000
  - INR>1.7
  - Partial Thromboplastin Time > 40 seconds
  - Systolic Blood Pressure > 185 mmHg
  - Diastolic Blood Pressure > 110 mmHg
  - Patient refusal

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Pharmacy
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary:

Evidence for this new eMeasure is the same as the existing measure #0437 STK 04: Thrombolytic Therapy:

- The body of evidence consistently supports that thrombolytic therapy results in a significant net reduction in the proportions of patients dead or dependent in activities of daily living due to ischemic stroke. Based on these findings, clinicians may choose to use thrombolytic therapy in selected patients within three hours of symptom onset.
- Class I, Level of Evidence A – 2007 AHA/ASA Guidelines for the Early Management of Adults with Ischemic Stroke, page 1676
  - Intravenous rtPA (0.9mg/kg, maximum dose 90 mg) is recommended for selected patients who may be treated within 3 hours of onset of ischemic stroke. Physicians should review the criteria outlined in Table 11 (which are modeled on those used in the NINDS trial) to determine the eligibility of the patient. A recommended regimen for observation and treatment of the patient is described in Table 12. This recommendation has not changed from previous statements.
- The 2012 Committee that reviewed the chart-abstracted version of this measure agreed that this measure was supported by strong evidence.

Guidance from the Evidence Algorithm:
Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency high(Box 5) → Box 5a High

Questions for the Committee:

- *The developer attests the underlying evidence for the measure #0437 (chart-abstracted version) has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the chart-abstracted measure has not changed and can be applied to the eMeasure, therefore there is no need for repeat discussion and voting on Evidence?*

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer states no performance data for the electronic version of this measure are yet available.
- This eCQM is currently included in the Hospital Inpatient Quality Reporting Program (IQR) and EHR Incentive program.
  - In CY2016, CMS is requiring organizations participating in HIQR to electronically submit one quarter of data on 4 of 28 available eCQMs. This measure is 1 of the 28 eCQMs.
- The developer also provided performance data from the chart abstracted measure which seems to be very high with little to no room for improvement.

Disparities:

- The developer does not provide disparities data from use of this measure.
- Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age."

Questions for the Committee:

- *Is there a gap in care that warrants a national performance eMeasure?*
- *How can this measure be used to better understand disparities in care and outcomes for certain groups?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

***1a. Evidence** to Support Measure Focus*

Comments: **No new studies/information

**Evidence is based upon the companion paper measure, which is high and does not need further discussion.

1b. Performance Gap

Comments: **performance data from chart abstracted measure provided; no disparities data provided

**The performance gap data from the paper measure does show opportunity for improvement. The literature also supports room for improvement. The measure does not provide data regarding detailed performance in disparate patient subgroups. It is an overall hospital measure. The emeasure will provide greater detail on reasons for exclusion and should capture better quality data in order to truly determine whether a performance gap remains in the future.

1c. High Priority (previously referred to as High Impact)
Comments: **NA
**N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic Clinical Data, Electronic Health Record (EHR), Imaging/Diagnostic Study, Laboratory Data, and Pharmacy Data. This is an eMeasure.

Specifications:

- HQMF specifications for eMeasure are included in the document set on SharePoint. See eMeasure Technical Review below.
- The developer states that changes have been made to the eCQM specifications to reflect revisions to the chart abstracted measure from which this measure is derived.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor review:

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Measure logic Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously Demonstrated through Bonnie testing
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible. This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment.

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☒ **No**

Method(s) of reliability testing:

- Dataset used for testing included 22 synthetic records created in the Bonnie testing system. Measure not tested in EHR.
- Data element validity testing was performed and will count for data element reliability as well – see validity testing section.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results from the Bonnie tool demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm: Precise specifications (Box 1) →empiric reliability testing as specified (Box 2 → empiric validity testing of patient-level data (Box 3) → YES: Use rating from validity testing of patient-level data elements (Box 10)→ Assessed the measure logic only (Box 11) →Low certainty or confidence data used in measure are valid (Box 12b)→Low (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Preliminary rating for reliability: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- The Bonnie testing tool and environment with 22 synthetic patient records were used to test the measure logic and value sets. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure.
- The numerator, denominator, exclusions/exceptions , and logic statement were tested to confirm actual results met expectations.
- The developer also conducted a review of the measure specifications to confirm the measure logic was properly expressed with the current version of the QDM and that the logic matches the clinical intent of the measure.

Validity testing results:

- The testing results from the Bonnie tool reached 100% coverage and confirmed there was a test case for each pathway of logic (negative and positive test cases).
- The measure also had a 100% passing rate which confirmed that all the test cases performed as expected.
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer submitted additional information on exclusions and meaningful differences.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Is it reasonable to assume that the impact of exclusions will be similar for the eMeasure and the chart-abstracted measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer submitted additional information on exclusions and meaningful differences.

Question for the Committee:

- *For meaningful differences in the chart-abstracted measure the developer refers to the performance scores over time. Does the SC think the eMeasure will demonstrate similar results to the chart-abstracted measure which show little to no room for improvement?*

2b6. Comparability of data sources/methods: Not Applicable

2b7. Missing Data

- The developer describes the handling of missing data but does not provide information on the frequency of missing data or potential impact on measure results due to limited testing abilities using the Bonnie tool.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → threats to validity empirically assessed to some extent (Box2) → empirical testing of data elements using BONNIE tool (Box 3,10) → Moderate

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **Specifications consistent

**Bonnie testing suggests that the validity is acceptable as there was 100% coverage among the 22 synthetic patient records, indicating that there was a test case for each logic pathway, and there was a 100% passing rate indicating that the test cases performed as expected.

2a2. Reliability Testing

Comments: **Data element testing only

2b2. Validity Testing

Comments: **Yes not assessed

**There does not appear to be sufficient missing data to suggest a validity threat.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Reliability testing only done at data element level not measure score level

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is EHR; all data elements are in defined fields in EHR.
- The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible.
- This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment.
- Per developer, they were unable to directly assess data accuracy of the data elements in an EHR system. Workflow or the degree to which the data elements are captured during the course of care was also not assessed.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **All data elements available in EHR, good feasibility

**The required elements are routinely generated during care delivery for stroke at stroke centers, and therefore utilization of this eMeasure at stroke centers should be feasible per this criteria, and the Bonnie testing further supports overall feasibility. However, as with reliability and validity, it would seem that evidence that this will be feasible in EHR systems from different providers would be critical to assessing feasibility for general use.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- CMS Hospital Inpatient Quality Reporting Program (IQR)/EHR Incentive Program: The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. The EHR Incentive Program provides incentive payments to promote the adoption and meaningful use of interoperable HIT and qualified EHRs.
- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results:

- The developer states that while the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are currently available.
- In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs, one of which is this measure. Data will be accepted through February 28th, 2017.

Unexpected findings (positive or negative) during implementation:

- Developer did not identify unexpected findings during implementation.

Potential harms:

- Developer did not identify any unintended consequences related to this measure.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Not publicly reported, but in CMS accountability program through 2017; benefits outweigh potential unintended consequences

**Usability is similar to the usability for the non-eMeasure upon which this is based.

Criterion 5: Related and Competing Measures

Related or competing measures:

- 0164 : Fibrinolytic Therapy received within 30 minutes of hospital arrival
- 0288 : Fibrinolytic Therapy Received Within 30 Minutes of ED Arrival
- 1952 : Time to Intravenous Thrombolytic Therapy
- 0437 : STK 04: Thrombolytic Therapy

- 0242 : Stroke and Stroke Rehabilitation: Tissue Plasminogen Activator (t-PA) Considered – AMA-PCPI – no longer NQF endorsed

Harmonization:

- Measures 0288 and 0164 are AMI (Acute Myocardial Infarction) measures and are harmonized to the extent that the measures utilize some of the same data elements.
- The target population for measure 1952 from the American Heart Association/American Stroke Association also includes patients hospitalized for acute ischemic stroke; however, the measure captures average door-to-needle time and uses a target of less than 60 minutes rather than the proportion of patients who arrive within 2 hours and receive t-PA within 3 hours of time last known well.
- Measure 0242 is a physician level measure with a targeted population of ischemic stroke patients identified through CPT codes and could extend to the outpatient setting.
- #0437 and #2834 are completely harmonized to the extent possible, given the fact that the data source for #0437 is the paper medical record, and the data source for #2834 is the electronic health record.

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0437 **NQF Project:** Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. **(evaluation criteria)**

1c.1 **Structure-Process-Outcome Relationship** *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The focus of the measure is to improve neurological outcomes for ischemic stroke patients through the timely initiation of intravenous (IV) thrombolytic therapy (t-PA) to carefully screened, eligible candidates.

IV thrombolytic therapy initiated >> improved neurological outcomes >> decreased morbidity and mortality.

1c.2-3 **Type of Evidence** *(Check all that apply)*:
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 **Directness of Evidence to the Specified Measure** *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:

This measure is consistent with the guidelines recommended by the American Heart Association/American Stroke Association for the early management of adults with ischemic stroke. The administration of thrombolytic agents to carefully screened, eligible patients with acute ischemic stroke has been shown to be beneficial in several clinical trials. These included two positive randomized controlled trials in the United States: The National Institute of Neurological Disorders and Stroke (NINDS) Studies, Part I and Part II. Based on the results of these studies, the Food and Drug Administration approved the use of intravenous recombinant tissue plasminogen activator (IV r-TPA or t-PA) for the treatment of acute ischemic stroke when given within 3 hours of stroke symptom onset. A large meta-analysis controlling for factors associated with stroke outcome confirmed the benefit of IV t-PA in patients treated within 3 hours of symptom onset. While controversy still exists among some specialists, the major society practice guidelines developed in the United States all recommend the use of IV t-PA for eligible patients. Physicians with experience and skill in stroke management and the interpretation of CT scans should supervise treatment.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: Relevant literature supporting thrombolysis for acute ischemic stroke was identified through a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in The Cochrane Library 2009, Issue 4. The search strategy employed the Cochrane Stroke Group Trials Register (last searched October 2008), MEDLINE (1966 to October 2008) and EMBASE (1980 to October 2008). Selection criteria involved randomized trials of any thrombolytic agent compared with control in patients with definite ischemic stroke (i.e., computerized tomography (CT) or magnetic resonance imaging (MRI) having excluded intracranial hemorrhage prior to randomization).

The review included 26 trials involving 7152 patients. This review included all new trials completed and made public since 2003, as well as additional data published since 2003 from trials included in earlier versions of the review. The total number of patients included in the review represented a 10-fold increase in the number of study patients since the first review in 1990. Very few of the patients (0.5%) were aged over 80 years.

The primary outcomes measured were death or dependency, as defined by modified Rankin score of 3 to 6, and death at the end of follow-up. Not all trials contributed data to each outcome. The trials tested urokinase, streptokinase, recombinant tissue plasminogen activator, recombinant pro-urokinase or desmoteplase, Four trials used intra-arterial administration, the rest used the intravenous route. Most data came from trials that started treatment up to six hours after stroke; three trials started treatment up to nine hours and one small trial up to 24 hours after stroke. About 55% of data (patients and trials) came from trials testing intravenous tissue plasminogen activator.

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting thrombolysis for acute ischemic stroke is high. The Cochrane Review noted above confirmed that thrombolytic treatment can reduce the risk of disability post-ischemic stroke, despite the bleeding risks. This review was conducted by two review authors who applied the selection criteria and extracted data. The reviewers assessed trial quality and verified the extracted data with the principal investigators of all major trials for both published and unpublished data. Specifically, the reviewers hand-searched journals from 1979 to April 1994; contacted 321 pharmaceutical companies for more information about trials known to exist and unknown based on prior searches; examined references quoted in thrombolytic therapy papers; contacted principal investigators of trials in Europe, North American, Japan, China , and Australia; and, attended multiple international conferences on stroke and thrombolysis since 1991). Every effort was made to identify truly randomized trials of thrombolytic therapy compared with placebo or open control in patients with acute ischemic stroke. Trials that were not truly random, such as dose-range-finding studies, were excluded.

According to the Cochrane reviewers, many trials had some imbalance in key prognostic variables. Several trials did not have complete blinding of outcome assessment. Thrombolytic therapy administered up to six hours after ischemic stroke significantly reduced the proportion of patients who were dead or dependent (modified Rankin 3 to 6) at three to six months after stroke (odds ratio (OR) 0.81, 95% Confidence Interval (CI) 0.73 to 0.90). Thrombolytic therapy increased the risk of symptomatic hemorrhage (OR 3.49, 95% CI 2.81 to 4.33) and death within three to six months after stroke (OR 1.31, 95% CI 1.14 to 1.50). Treatment within three hours of stroke appeared more effective in reducing death or dependency (OR 0.71, 95% CI 0.52 to 0.96) with no statistically significant adverse effect on death (OR 1.13, 95% CI 0.86 to 1.48). There was heterogeneity among the trials in part attributable to concomitant antithrombotic drug use (P =0.02), stroke severity and time to treatment. Antithrombotic drugs given soon after thrombolysis may increase the risk of death.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of

evidence consistently supports that thrombolytic therapy results in a significant net reduction in the proportions of patients dead or dependent in activities of daily living due to ischemic stroke. Based on these findings, clinicians may choose to use thrombolytic therapy in selected patients within three hours of symptom onset.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)***:**
There is strong evidence on the immediate hazards and net benefits of thrombolytic therapy in the treatment of acute ischemic stroke. Thrombolytic therapy administered up to six hours after ischemic stroke significantly reduced the proportion of patients who were dead or dependent (modified Rankin 3 to 6) at three to six months after stroke (odds ratio (OR) 0.81, 95% Confidence Interval (CI) 0.73 to 0.90). Treatment within three hours of stroke appeared more effective in reducing death or dependency (OR 0.71, 95% CI 0.52 to 0.96) with no statistically significant adverse effect or death (OR 1.13, 95% CI 0.86 to 1.48).

Treating ischemic stroke patients with IV t-PA within 3 hours of time last known well improves functional outcomes at 3 months. Considering the mean lifetime cost of ischemic stroke, estimated at $140,048 per person, thrombolytic therapy is also cost-effective. Fagan et al. (1998) used data from the NINDS rt-PA Stroke Trial and developed a Markov model to compare costs per 1,000 eligible t-PA patients with the costs per 1,000 untreated patients. This analysis revealed a greater than 90% probability of costs savings when eligible patients are treated with t-PA within 3 hours. In the NINDS rt-PA Stroke Trial, the average length of stay was significantly shorter for patients treated with t-PA than the placebo group (10.9 versus 12.4 days; p=0.002), and more patients were discharged to home than to an inpatient rehabilitation facility or a nursing home (48% versus 36%; p=0.002). The Markov model estimated an increase in hospitalization costs of $1.7 million and a decrease in rehabilitation costs of $1.4 million and nursing home cost of $4.8 million per 1,0000 eligible t-PA patients. Furthermore, the analysis estimated the long-term impact on health outcomes to be 564 quality-adjusted life-years saved over 30 years per 1,000 patients.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: American Heart Association. During our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, and provided information that was accessible and met the requirements set out in this measure maintenance form.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.13 Grade Assigned to the Body of Evidence: Class I, Level of Evidence A

1c.14 Summary of Controversy/Contradictory Evidence: The benefit of IV thrombolytic therapy within 3 hours of stroke symptom onset is undisputed. No position against the administration of thrombolytic therapy to select, eligible ischemic stroke patients was noted in the literature. However, there is still not enough evidence to answer some questions. More randomized trials are needed to determine the full impact of benefit, the latest time window for safe administration, the cut-off age for treatment, the type of stroke and grades of severity most likely to respond favorably to treatment, and the impact of various co-morbidities and concomitant drug therapies on outcome.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- del Zoppo GJ, Saver JL, Jauch EC, Adams HP. Expansion of the Time Window for Treatment of Acute Ischemic Stroke With Intravenous Tissue Plasminogen Activator: A Science Advisory From the American Heart Association/ American Stroke Association. Stroke. 2009;40:2945-2948.
- Diagnosis and Initial Treatment of Ischemic Stroke, Institute for Clinical Systems Improvement (ICSI), 2001.
- Fagan SC, Morgenstern LB, Petitta A, Ward RE, Tilley BC, Marler JR, Levine SR, Broderick JP, Kwiatkowski TG, Frankel M, Brott TG, Walker MD, and The NINDS rt-PA Stroke Study Group. Cost-efffectiveness of tissue plasminogen activator for acute ischemic stroke. Neurology. 1998;50(4):883-890.
- Guyatt GH, Akl EA, Crowther M, Gutterman DD, Schu?nemann HJ, and for the American College of Chest Physicians Antithrombotic Therapy and Prevention of Thrombosis Panel. Executive summary: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141: 33S.
- Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Gidetti D, et. al. Thrombolysis with Alteplase 3 to 4.5 hours after acute ischemic stroke. The European Cooperative Acute Stroke Study (ECASS) Investigators. NEJM. 2008;359(13):1317-29.

- Hacke W, Kaste M, Fieschi C, et al. Intravenous thrombolysis with recombinant tissue plasminogen activator for acute hemispheric stroke. The European Cooperative Acute Stroke Study (ECASS). JAMA 1995;274:1017-1025.
- Management of Patients with Stroke. Assessment, investigation, immediate management and secondary prevention, Scottish Intercollegiate Guidelines Network, 1997.
- Marler JR, Tilley BC, Lu M, Brott TG, Lyden PC, Grotta JC, Broderick JP, Levine SR, Frankel MP, Horowitz SH, Haley EC, Lewandowski CA, Kwiatkowski TP. Early stroke treatment associated with better outcome The NINDS rt-PA Stroke Study. Neurology 2000;55: 1649-1655.
- Sacco RL, Adams R, Albers G, Alberts MJ, Benavente O, Furie K, Goldstein LB, Gorelick P, Halperin J, Harbaugh R, Johnston SC, Katzan I, Kelly-Hayes M, Kenton EJ, Marks M, Schwamm LH, Tomsick T. Guidelines for Prevention of Stroke in Patients With Ischemic Stroke or Transient Ischemic Attack: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association Council on Stroke: Co-Sponsored by the Council on Cardiovascular Radiology and Intervention. Stroke. Vol. 37, 2006:577.
- STROKE the First Hours Guidelines for Acute Treatment, National Stroke Association, 2000.
- The ATLANTIS, ECASS, and NINDS rt-PA Study Group Investigators. Association of Outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke Trials. Lancet 2004;363:768-774.
- The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. New England Journal of Medicine 1995;333:1581-1587.
- Wardlaw JM, Murray V, Berge E, del Zoppo GJ for The Cochrane Collaboration. Thrombolysis for acue ischaemic stroke (review). The Cochrane Library. 2009:4: 1-131.

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)*:
2007 AHA/ASA Guidelines for the Early Management of Adults with Ischemic Stroke, page 1676.

Class I, Level of Evidence A
Intravenous rtPA (0.9mg/kg, maximum dose 90 mg) is recommended for selected patients who may be treated within 3 hours of onset of ischemic stroke. Physicians should review the criteria outlined in Table 11 (which are modeled on those used in the NINDS trial) to determine the eligibility of the patient. A recommended regimen for observation and treatment of the patient is described in Table 12. This recommendation has not changed from previous statements.

1c.17 Clinical Practice Guideline Citation: Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A Guideline From the American Heart Association/American Stroke, Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1674-1677.

1c.18 National Guideline Clearinghouse or other URL:
http://www.guidelines.gov/content.aspx?id=10911&search=early+management+of+adults+with+ischemic+stroke

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: This guideline was approved by the American Heart Association Science Advisory and Coordinating Committee on January 6, 2007. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.23 Grade Assigned to the Recommendation: Class I, Level of Evidence A

1c.24 Rationale for Using this Guideline Over Others: The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of

stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: High **1c.26** Quality: High**1c.27** Consistency**:** High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0437_Evidence_MSF5.0_Data-635827722587031131.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Studies have demonstrated that the early administration of thrombolytic therapy within 3 hours of stroke symptom onset can significantly improve neurologic outcomes at 3 months in patients with ischemic stroke. The European Cooperative Acute Stroke Study (ECASS) III trial indicated that intravenous rtPA can be given safely to, and can improve outcomes for, carefully selected patients treated 3 to 4.5 hours after stroke; however, as the NINDS investigators concluded, the earlier that IV thrombolytic therapy is initiated, the better the patient outcome. Despite strong recommendations from the American Academy of Neurology, the American Heart Association, and the American College of Chest Surgeons, thrombolytic therapy is used in only a small proportion of ischemic stroke patients overall and in only a minority of eligible candidates.

Healthcare organizations that track IV thrombolytic administration for internal quality improvement purposes have seen a significant increase in the measure rate over time. The chart-abstracted measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This measure is a legacy eCQM that is currently included in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data for the electronic version of the measure are yet available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs. This measure is one of the 28.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
Many patients with acute ischemic stroke arrive at the hospital within three hours of stroke onset without documented contraindications who still do not receive intravenous (IV) thrombolysis (t-PA). Although rates of IV t-PA administration have

improved over time, there is still less than optimal performance especially in the lower quartile and decile of hospitals. In October, 2009, The Joint Commission added the chart-abstracted stroke (STK) measure set as a new core measure option to meet performance measure requirements for Joint Commission hospital accreditation purposes. Joint Commission ORYX performance measure data for 4Q 2009 yielded an average measure rate of 48.9% from 39 hospitals collecting data for this measure (n=233 patients). The average rate for all hospitals collecting data for this measure (i.e., 1099 hospitals; n=14,907 patients) is currently 84.5%, indicating that a potential performance gap of 15% persists if the optimal rate is 100%. Below is the specified level of analysis for STK-4 beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 233 denominator cases; 114 numerator cases; 39 hospitals; 0.48927 national aggregate rate; 0.54417 mean of hospital rates; 0.41298 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.625 50th percentile rate/median rate; 0.03704 25th percentile rate/lower quartile; and, zero 10th percentile rate.

CY 2010: 1728 denominator cases; 1053 numerator cases; 129 hospitals; 0.60938 national aggregate rate; 0.56977 mean of hospital rates; 0.37503 standard deviation; 1.0 90th percentile rate; 0.9 75th percentile rate/upper quartile; 0.66667 50th percentile rate/median rate; 0.21212 25th percentile rate/lower quartile; and, zero 10th percentile rate.

CY 2011: 1942 denominator cases; 1333 numerator cases; 136 hospitals; 0.68641 national aggregate rate; 0.57338 mean of hospital rates; 0.37267 standard deviation; 1.0 90th percentile rate; 0.89737 75th percentile rate/upper quartile; 0.71964 50th percentile rate/median rate; 0.23611 25th percentile rate/lower quartile; and, zero 10th percentile rate.

CY 2012: 1807 denominator cases; 1393 numerator cases; 136 hospitals; 0.77089 national aggregate rate; 0.67283 mean of hospital rates; 0.34402 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.7735 50th percentile rate/median rate; 0.5 25th percentile rate/lower quartile; and, zero 10th percentile rate.

CY 2013: 3135 denominator cases; 2493 numerator cases; 224 hospitals; 0.79522 national aggregate rate; 0.74944 mean of hospital rates; 0.3186 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.875 50th percentile rate/median rate; 0.66667 25th percentile rate/lower quartile; and, zero 10th percentile rate.

CY 2014: 14,907 denominator cases; 12,598 numerator cases; 1099 hospitals; 0.84511 national aggregate rate; 0.75482 mean of hospital rates; 0.32814 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.88889 50th percentile rate/median rate; 0.66667 25th percentile rate/lower quartile; and, zero 10th percentile rate.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
Since the last endorsement period, several large studies have evaluated trends in thrombolysis utilization across the United States. While disparities still exist for many groups, the disparity gap has narrowed for certain groups, e.g., older adults, young adults, women, and young blacks, since the last submission.

George and colleagues (2015) analyzed discharge data from 2005 to 2010 in the Nationwide Inpatient Sample (NIS) of the Agency for Healthcare Research and Quality (AHRQ), Healthcare Cost and Utilization Project (Rockville, MD). This analysis identified persistent racial disparities for blacks and Hispanics consistent with those found previously for blacks and Hispanics (Kimball, et al. 2012), as well as, a trend to treat older adults who would not have been considered t-PA candidates in earlier years and increased utilization rates for rural patients. From this database, 512,429 older adults (> 65 years of age) with acute ischemic stroke who received an injection or infusion of a thrombolytic agent were identified and divided into three age groupings (65-74, 75-84, > 85 years). Among U.S. hospitals with acute stroke patients, there were an estimated 1462 hospitals (32%) administering thrombolysis in 2005 and 1884 hospitals (43%) in 2010 (trend P = 0.003). For older adult stroke admissions, the rate of thrombolysis increased from 1.7% (95% CI: 1.6-1.8%) in 2005 to 5.4% (95% CI: 5.2-5.6%) in 2010, representing a three-fold increase in thrombolysis rates for older adults (trend P < 0.001). The largest increases occurred for individuals > 85 years of age with an approximate four-fold rate increase from 2005 to 2010. Rates of administration increased three-fold for urban patients and urban hospitals. Thrombolysis rates also increased for rural patients (0.9% in 2005 vs. 3.3% in 2010; trend P < 0.001), and rural hospitals increased at a slower rate (0.5% in 2005 vs. 1.7% in 2010; (trend P < 0.001). Low volume hospitals increased their rates of thrombolysis in older adults to a lesser degree than higher

volume centers.

A second study using discharge data obtained from the Nationwide Inpatient Sample between 2001 and 2009 reported that disparities for young blacks has significantly improved in recent years (Kansara, et al. 2013). Between 2001 and 2009, there were an estimated 4,917, 217 admissions for acute ischemic stroke. Of these, 204,703 (4.16%) were young patients with a mean age approximately 37 years. The use of thrombolysis for young acute ischemic stroke patients increased 270% during this time period. The increased rate was noted across all races, including white, black, and nonwhite/nonblack populations. Unlike previous studies that reported that black patients were less likely to receive thrombolysis, this study found that a greater percentage of young black patients with acute ischemic stroke (5.45%) received thrombolysis that young white patients with acute ischemic stroke (4.57%) in 2009.

A univariate analysis of more than one million (N=1,093,895) acute ischemic stroke patients from 1683 hospitals participating in the American Heart Association's Get With the Guidelines-Stroke database was conducted to evaluate changes in the patterns of IV t-PA use over the 9-year period from April 2003 to December 2011. IV thrombolytic use has changed over time with a broader range of patients treated in later years. According to this analysis, the proportion of patients age > 85 years treated with IV t-PA increased from 10.5% in 2003-2005 to 16.4% in 2010-2011 (P<0.001). Also, the gender distribution of t-PA use changed slightly, with the proportion of t-PA use among women increasing from 48.6% t0 51%. The population receiving t-PA also became more diverse, with nonwhites accounting for 21.1% of t-PA use in 2003 to 2005 but 28.9% in 2010 to 2011(P<0.001) (Schwamm, et al. 2013). This study is among the first to describe the temporal trends of the past decade in IV t-PA use in patients with acute ischemic stroke in a clinically derived dataset from a sizeable cohort of U.S. hospitals nationwide.

Another published study utilized the American Heart Association's Get With the Guidelines-Stroke registry linked with Medicare claims data set to examine whether 30-day and 1-year outcomes differed by race/ethnicity among older patients with acute ischemic stroke (Qian, et al., 2013). Compared with other race/ethnicity groups, non-Hispanic black patients were less likely to receive IV t-PA in less than 3 hours from stroke onset. Relative to whites, black and Hispanic patients had higher adjusted 1-year all-cause rehospitalization (black: adjusted odds ratio, 1.28 [95% CI, 1.21-1.37]; Hispanics: adjusted odds ratio 1.22 [95% CI, 1.11-1.35]. Non-Hispanic black patients were more likely to be treated at high-volume and academic hospitals, which were generally located in the south. These findings were based on an analysis of 200,900 patients, including 20514 non-Hispanic blacks (10.2%), with acute ischemic stroke greater than 65 years of age from 926 U.S. centers participating in the GWTG-Stroke program from April 2003 through December 2008.

A prior 2011 report from the American Heart Association/American Stroke Association reported that racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al., 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

In the national REGARDS cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted,

age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years of age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age: RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer,et al., 2010).

In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).

Minorities are less likely to use emergency medical services and more likely to wait longer before going to the hospital than are non-Hispanic whites. After hospital arrival, blacks or African Americans and possibly Hispanics experience longer wait times in the emergency department which results in treatment delays that include administration of thrombolytic therapy. Although unmeasured factors may play a role in these delays, the presence of bias in the delivery of care cannot be excluded (Cruz-Flores, et al., 2011).

The Department of Neurology and Stroke Program, Wayne State University School of Medicine, Detroit, MI (Bhattacharya P, et al., 2011) explored racial differences in the delivery of care to patients with acute stroke between hospitals certified in primary stroke care by The Joint Commission and noncertified hospitals. A retrospective chart review of 574 patients (25.1% African American) with ischemic stroke admitted to five Joint Commission certified primary stroke centers and five non-certified hospitals was conducted. Similar to previous studies, Bhattacharya found that African Americans often did not receive intravenous tPA because of a delay in hospital arrival. Whites were more likely to arrive by emergency transport services (65.5% vs 51.1%; P= 0.004) to be evaluated by a stroke team (19.1% vs. 7.7%; P=0.001), and to have documented National Institutes of Health Stroke Scale (NIHSS) score (40.2% vs. 29/9%; P=0.03); however, the number of white and black patients who received IV t-PA was not statistically different (2.1% in African Americans, 3.5% in Caucasians; P=0.40).

A larger study of 1044 patients (74% African American, 19% non-Hispanic white) with ischemic stroke (Hsia AW, et al., 2011), found that blacks were one-third less likely than whites to receive IV t-PA (3% vs.10%, P<0.001). Blacks were less likely than whites to present in 3 hours of symptom onset (13% vs. 21%; P=0.004). They were also less likely to be eligible candidates for thrombolytic therapy (5% vs. 13%; P<0.001). Of those patients who presented in 3 hours, blacks were almost half as likely to be treated with IV t-PA when compared to whites (27% vs. 46%; P=0.023).

A recent study from the Mayo Clinic, Rochester, MN (Naser DM, et al., 2011) investigated possible racial and ethnic disparities in the administration and outcome of recombinant tissue plasminogen activator (rt-PA) therapy for acute ischemic stroke in whites, blacks, Hispanics, and Asian/Pacific Islanders. Patients with a primary diagnosis of acute ischemic stroke who received rt-PA were identified using data from the National Inpatient Sample for 2001-2008 and stratified by race. The investigators analyzed the association of patient race on rt-PA utilization rate, in-hospital morbidity (i.e., discharges to a long-term care facility), intracranial hemorrhage (ICH) rate, and in-hospital mortality. Multivariate logistic regression analysis was performed to identify independent predictors of poor outcomes. Naser and colleagues concluded that whites had a higher rate of t-PA utilization than black and Hispanic patients (2.3% vs. 2.2% P= 0.07), although not statistically significant. Multivariate analysis of morbidity, mortality and ICH rates found that Asian/Pacific Islanders had significantly higher rates of mortality (odds ratio, 1.22, 95% CI, 1.91-2.11; P < .0001) compared with

whites. Thrombolytic utilization was greater in white and Asian/Pacific Islander patients than in black and Hispanic patients. Asian/Pacific Islander race was associated with increased risk of ICH and mortality with rt-PA administration.

• Bhattacharya P, Mada F, Salowich-Palm L, et al. Department of neurology and Stroke Program, Wayne State University School of Medicine, Detroit, Michigan. Are racial disparities in stroke care still prevalent in certified stroke centers? J Stroke Cerebrovasc Dis. 2011.
• Brown AF, Liang LJ, Vassar SD, et al. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-8.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, et al, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• George BP, Asemota AO, Dorsey ER, Haider AH, Smart BJ, Urrutia VC, Schneider EB. United States trends in thrombolysis for older adults with acute ischemic stroke. Clinical Neurology and Neurosurgery. 2015;139:16-23.
• Gillum RF, Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 2011;42(11):3294-6.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-75.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Hsia AW, Edwards DF, Morgenstern LB, et al. Racial disparities in tissue plasminogen activator treatment rate for stroke: a population-based study. Stroke. 2011;42(8):2217-21.
• Jacobs BS, Birbeck G, Mullard AJ, Hickenbottom S, Kothari R, Roberts S, Reeves MJ. Quality of hospital care in African Americans and white patients with ischemic stroke and TIA. Neurology.2006;66:809-14.
• Johnston SC, Fung LH, Gillum LA, Smith WS, Brass LM, Lichtman JH, Brown AN. Utilization of intravenous tissue-type plasminogen activator for ischemic stroke at academic medical centers.: the influence of ethnicity. Stroke. 2001; 32:1061-68.
• Kansara A, Chaturvedi S, Bhattacharya P. Thrombolysis and outcome of young stroke patients over the last decade: insights from the nationwide inpatient sample. J. Stroke Cerebrovasc Dis. 2013;6:799-804.
• Karve S, Balkrishnan R, Seiber E Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Morris DL, Rosamond W, Madden K, Schultz C, Hamilton S. Pre-hospital and emergency department delay after acute stroke: the Genetech Stroke Presentation Survey. Stroke. 2000;31:2585-90.
• Kimball MM, Neal D, Waters MF, Hoh BL, Race andincome disparity in ischemic stroke care: nationwide inpatient sample database, 2002-2008. Cerebrovasc Dis. 2012.
• Kissela BM, Khoury JC, Alwell K, et al. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, et al. Stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Nasar DM, Brinjikji W, Cloft HJ, Rabinstein AA. Racial and ethnic disparities in the use of intravenous recombinant tissue plasminogen activator and outcomes for acute ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, et al. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e-78-e82.
• Schwamm LH, Syed FA, Reeves MJ, Smith EE, Saver JL, Messe S, Bhatt DL, Grau-Sepulveda MV, Peterson ED, Fonarow GC. Temporal trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With the Guidelines-Stroke hospitals. Circ Cardiovasc Quality Outcomes. 2013;6:543-549.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality, Severity of illness

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States remains attributable to stroke. More women than men die of stroke each year. Women account for almost 60% of US stroke deaths. (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, et al., 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 ($P<0.05$) (US Burden of Disease Collaborators, 2013). Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

Thrombolytic therapy is one of the most promising treatments for acute ischemic stroke. The majority of strokes are due to blockage of an artery in the brain by a blood clot. Prompt treatment with clot dissolving (thrombolytic) drugs can restore blood flow before major brain damage has occurred. In the United States, Canada, and most European countries, alteplase/recombinant tissue plasminogen activator (rt-PA) has been approved for use within three hours of stroke symptom onset. Successful treatment is likely to improve neurological outcomes for ischemic stroke patients at three months and later; however, intracranial hemorrhage is a serious complication of therapy and may be fatal (Jauch EC, et al., 2013).

Clinical practice guidelines for intravenous thrombolysis with rt-PA (Adam HP, et al., 2007) cite the National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Study (1995), which partially supported the approval of rt-PA by the U.S. Food and Drug Administration (FDA). The NINDS trial was conducted in two consecutive parts. Part A trials in the 1980s studied very low doses of thrombolytic therapy, given daily by intravenous route for several days. The onset of treatment occurred anytime from 5 to 14 days post symptom onset. Part A trials did not collect data on functional outcome. Part B trials from the 1990s and later used a single large dose of thrombolytic drug (80 to 100 mg rt-PA), given intravenously (IV) or intra-arterially (IA) within three, six, nine, or 24 hours of stroke. The primary end point in Part B of the study was a favorable outcome, defined as complete or nearly complete neurological recovery 3 months after stroke. Favorable outcomes were achieved in 31% to 50% of patients treated with rt-PA, as compared to 20% to 38% of patients given placebo (Kwiatkowski TJ, et al., 1999). The benefit was similar one year after stoke. The major risk of treatment was symptomatic intracranial hemorrhage which occurred in 6.4% of patients treated with rt-PA and 0.6% of patients given placebo (Marler JR, et al., 2007).

In 2008, European Cooperative Acute Stoke Study (ECASS)-3, a multi-center, prospective, randomized, placebo-controlled trial, studied the administration of rt-PA between three and 4.5 hours of stroke symptom onset (Hacke W, et al., 2008). The trial enrolled 418 patients treated with rt-PA per the current dosing guidelines (i.e., 0.9 mg/kg (maximum of 90 mg) with 10% given as an initial IV bolus and the remainder infused over one hour) and compared them with 403 who were given placebo. The frequency of the primary efficacy outcome (i.e., modified Rankin Scale score of 0 to 1 at 90 days after treatment) was significantly greater with rt-PA (52.4%) than with placebo (45.2%; OR 1.34, 95% CI 1.02 to 1.76; risk ratio 1.16, 95% CI 1.01 to 1.34; P= 0.04). The point estimate for the degree of benefit seen in ECASS-3 (OR for global favorable outcome, 1.28, 95% CI 1.00 to 1.65) was less than the point estimate of benefit found in the pool of patients enrolled for 0 to 3 hours after stroke symptom onset in the NINDS study (OR 1.9, 95% CI 1.2 to 2.9).

Currently, researchers, along with the field, continue to debate the risk of intracerebral hemorrhage with IV rt-PA in certain patient

populations; however, the benefit of improved functional outcomes, and potential improvements in quality of life outweighs the decision to with hold treatment of the ischemia (Saposnick, 2012).

Although the expanded timeframe of 3 to 4.5 hours for thrombolytic therapy was found to be effective for ischemic stroke patients and without significant increase in hemorrhagic events, a treatment target of 3 hours remains the accepted recommendation as studies have shown the potential opportunity for improved outcomes is greater with earlier treatment. Delays in evaluation and initiation of therapy for eligible patients with acute ischemic stroke should be avoided.

1c.4. Citations for data demonstrating high priority provided in 1a.3

• Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A Guideline From the American Heart Association/American Stroke, Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1655-1711.
• Adams H, Adams R, Del Zoppo G, Goldstein LB. American Heart Association/American Stroke Association Guidelines Update A Scientific Statement From the Stroke Council of the Guidelines for the Early Management of Patients With Ischemic Stroke: 2005, Stroke.2005;36;916-923.
• Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• del Zoppo GJ, Saver JL, Jauch EC, Adams HP. Expansion of the Time Window for Treatment of Acute Ischemic Stroke With Intravenous Tissue Plasminogen Activator: A Science Advisory From the American Heart Association/ American Stroke Association. Stroke. 2009;40:2945-2948.
• George M., Xin Tong, McGruder H., Yoon P., Rosamond W., Winquist A., Hinchey J., Wall H., Pandey D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Gidetti D, et. al. Thrombolysis with Alteplase 3 to 4.5 hours after acute ischemic stroke. The European Cooperative Acute Stroke Study (ECASS) Investigators. NEJM. 2008;359(13):1317-29.
• Jauch EC, Saver JL, Adams HP, Bruno A, Connors JJ, Demaerschalk BM, Khatri P, McMullan PW, Qureshi AI, Rosenfield K, Scott PA, summers DR, Wang DZ, Wintermark M, Yonas, H. Guidelines for the early management of patients with acute ischemic stroke: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44:893-899.
• Kwiatkowski TG, Libman RB, Frankel M, Tilley BC, Morgenstern LB, Lu M, Broderick JP, Lewandowski CA, Marler HR, Levine SR, Brott T. National Institute of Neurological Disorders and Stroke Recombinant Tissue Plasminogen Activator Study Group. Effects of tissue plasminogen activator for acute ischemic stroke at one year. NEJM. 1999;340:1781-87.
• Marler JR, Tilley BC, Lu M, Brott TG, Lyden PC, Grotta JC, Broderick JP, Levine SR, Frankel MP, Horowitz SH, Haley EC, Lewandowski CA, Kwiatkowski TP. Early stroke treatment associated with better outcome The NINDS rt-PA Stroke Study. Neurology 2000;55:1649-1655.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e-78-82.
• Saposnik G, Fang J, Kapral MK, Tu JV, Mamdani M, Austin P, Johnston C. The iScore Predicts Effectiveness of Thrombolytic Therapy for Acute Ischemic Stroke. Stroke. 2012;43:1315-1322.
• The ATLANTIS, ECASS, and NINDS rt-PA Study Group Investigators. Association of Outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke Trials. Lancet 2004;363:768-774.
• The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. New England Journal of Medicine 1995;333:1581-1587.
• US Burden of Disease Collaborators. The state of US health, 1990–2010: burden of diseases, injuries, and risk factors. JAMA. 2013;310:591–608.

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input*

was obtained.)
Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Health and Functional Status : Functional Status, Prevention, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/ecqm_library.html

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is an eMeasure **Attachment:** STK4_MAT.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** STK4_v5_Wed_Apr_01_12.15.32_CDT_2015.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Changes have been made to the eCQM specifications in order to reflect revisions to the chart abstracted measure from which this measure is derived.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Acute ischemic stroke patients for whom IV thrombolytic therapy was initiated at this hospital within 3 hours (less than or equal to 180 minutes) of when it was witnessed or reported that the patient was last known to be without the signs and symptoms of the current stroke or at his or her baseline state of health.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Thrombolytic (t-PA) Therapy Administration
- Thrombolytic (t-PA) Therapy is represented with the QDM datatype and value set of Medication, Administered: Thrombolytic (t-PA) Therapy (OID: 2.16.840.1.113883.3.117.1.7.1.226)

Baseline State
- Baseline State is represented with the QDM datatype and value set of Physical Exam, Performed: Baseline State (OID: 2.16.840.1.113883.3.117.1.7.1.417)

Emergency Department Visit
- Emergency Department Visit is represented with the QDM datatype and value set of Encounter, Performed: Emergency Department Visit (OID: 2.16.840.1.113883.3.117.1.7.1.292)

Time of Symptom Onset
- Time of Symptom Onset is represented with the QDM datatype and value set of Physical Exam, Performed: Time of Symptom Onset (OID: 2.16.840.1.113762.1.4.1045.14)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
Ischemic stroke patients admitted through the Emergency Department whose time of arrival is within 2 hours (less than or equal to 120 minutes) of the 1) time they were known to be at their baseline state of health; or 2) time of symptom onset if time last known at baseline state is not known.

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Principal Diagnosis of Ischemic Stroke
- Ischemic Stroke is represented with the QDM datatype and value set of Diagnosis, Active: Ischemic Stroke (OID: 2.16.840.1.113883.3.117.1.7.1.247)
- Ordinality: Principal (OID: 2.16.840.1.113883.3.117.1.7.1.14)

Non-Elective Inpatient Encounter
- Non-Elective Inpatient Encounter is represented with the QDM data type and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

Emergency Department Visit
- Emergency Department Visit is represented with the QDM datatype and value set of Encounter, Performed: Emergency Department Visit (OID: 2.16.840.1.113883.3.117.1.7.1.292)

Baseline State
- Baseline State is represented with the QDM datatype and value set of Physical Exam, Performed: Baseline State (OID: 2.16.840.1.113883.3.117.1.7.1.417)

Time of Symptom Onset
- Time of Symptom Onset is represented with the QDM datatype and value set of Physical Exam, Performed: Time of Symptom Onset (OID: 2.16.840.1.113762.1.4.1045.14)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Denominator Exclusions:
None.

Denominator Exceptions:

- Patients with comfort measures documented on the day of or the day after arrival
- Patients with intra-venous or intra-arterial Thrombolytic (t-PA) Therapy prior to arrival
- Patients with documentation of a National Institutes for Health Stroke Scale (NIHSS) score of zero in the emergency department
- Patients with Medical Reasons for not initiating IV thrombolytics documented by a physician/APN/PA or pharmacist on the day of or the day after arrival
- Patients with any of the following results within 180 minutes of the 1) time they were known to be at their baseline state of health; or 2) time of symptom onset:
  - Prothrombin Time > 15 seconds
  - Platelet Count <100,000
  - INR>1.7
  - Partial Thromboplastin Time > 40 seconds
  - Systolic Blood Pressure > 185 mmHg
  - Diastolic Blood Pressure > 110 mmHg
  - Patient refusal

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

Comfort Measures

- Comfort Measures are represented with the QDM datatypes and value set of:
- Intervention, Order: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)
- Intervention, Performed: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)

Reasons for Not Initiating IV Thrombolytic

- Medical Reason For Not Initiating IV Thrombolytic is represented with either of the following QDM datatypes and value sets:
  - Medication, Order not done: Medical Reason (OID: 2.16.840.1.113883.3.117.1.7.1.473)
  - Medication, Administered not done: Medical Reason (OID: 2.16.840.1.113883.3.117.1.7.1.473)
  - Medication, Order: t-PA ingredient specific (OID: 2.16.840.1.113762.1.4.1021.6)
  - Medication, Administered: Thrombolytic (t-PA) Therapy (OID: 2.16.840.1.113883.3.117.1.7.1.226)
- Laboratory Test, Performed: Prothrombin Time (OID: 2.16.840.1.113762.1.4.1045.24)
- Laboratory Test, Performed: Platelet Count (OID: 2.16.840.1.113883.3.117.1.7.1.267)
- Laboratory Test, Performed: INR (OID: 2.16.840.1.113883.3.117.1.7.1.213)
- Laboratory Test, Performed: Partial Thromboplastin Time (OID: 2.16.840.1.113762.1.4.1045.25)
- Physical Exam, Performed: Systolic Blood Pressure (OID: 2.16.840.1.113883.3.526.2.1044)
- Physical Exam, Performed: Diastolic Blood Pressure (OID: 2.16.840.1.113883.3.526.2.1045)
- Medication, Administered not done: Patient Refusal (OID: 2.16.840.1.113883.3.117.1.7.1.93)
- Medication, Order not done: Patient Refusal (OID: 2.16.840.1.113883.3.117.1.7.1.93)

Thrombolytic (t-PA) Therapy Already Administered

- Thrombolytic (t-PA) Therapy is represented with either of the following QDM datatypes and value sets:
  - Medication, Administered: Thrombolytic (t-PA) Therapy (OID: 2.16.840.1.113883.3.117.1.7.1.226)
  - Procedure, Performed: Intravenous Thrombolytic (t-PA) Therapy (OID: 2.16.840.1.113762.1.4.1045.15)

National Institutes for Health Stroke Scale (NIHSS)

- National Institutes for Health Stroke Scale (NIHSS) is represented with the QDM datatype and value set Risk Category Assessment: National Institute of Health Stroke Scale (OID: 2.16.840.1.113883.3.117.1.7.1.269)

Baseline State

- Baseline State is represented with the QDM datatype and value set of Physical Exam, Performed: Baseline State (OID: 2.16.840.1.113883.3.117.1.7.1.417)

Time of Symptom Onset

- Time of Symptom Onset is represented with the QDM datatype and value set of Physical Exam, Performed: Time of Symptom Onset (OID: 2.16.840.1.113762.1.4.1045.14)

Non-Elective Inpatient Encounter

- Non-Elective Inpatient Encounter is represented with the QDM data type and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

Emergency Department Visit
- Emergency Department Visit is represented with the QDM datatype and value set of Encounter, Performed: Emergency Department Visit (OID: 2.16.840.1.113883.3.117.1.7.1.292)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable, the measure is not stratified.

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable.

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. Type of score:
Rate/proportion
If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. Calculation Algorithm/Measure Logic *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
See attached HQMF file.

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable.

S.21. Survey/Patient-reported data *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*

IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
eMeasures are calculated using only the structured data collected in certified EHR technology (CEHRT). Data not present in the structured field from which the measure draws will not be included in the measure calculation.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Pharmacy

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Hospitals report EHR data using Certified Electronic Health Record Technology (CEHRT), and by submitting Quality Reporting Document Architecture Category 1 (QRDA-1).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF0437_CMS91v4_STK4_Bonnie_Testing.xlsx,STK4_eCQM_testing_attachment-635907800145590291.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2834
Measure Title: STK04: Thrombolytic Therapy
Date of Submission: 1/14/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	X Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For <u>all</u> measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For <u>outcome and resource use</u> measures**, section **2b4** also must be completed.
- If specified for **<u>multiple data sources/sets of specificaitons</u>** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to <u>all</u> questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the sub criteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

<u>Note</u>: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or

whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒other: Bonnie Test Cases	☒ other: Bonnie Test Cases (see question 1.2)

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

This measure is a legacy eCQM (an NQF endorsed measure that has been re-specified into an eMeasure and is currently used in a federal quality program/programs). In accordance with the modified NQF requirements for testing of eMeasures, the Bonnie testing tool was used to simulate a testing environment where measure specifications and HQMF output are tested against synthetic test data. Measure developers rely on the results in Bonnie to confirm whether the measure logic is performing as expected. Reference the eCQI Resource Center website (https://ecqi.healthit.gov/ecqm-tools/tool-library/bonnie) or the Bonnie testing tool website (https://bonnie.healthit.gov/) for more information about Bonnie functionality and its role in measure development. Please also reference the Bonnie testing worksheet attachment for detailed Bonnie test cases and testing results for this measure.

While Bonnie does consume HQMF specifications, we are not able to test the measure with data from HQMF implemented in EHRs.

1.3. What are the dates of the data used in testing? The measure specifications were tested in the Bonnie testing environment which mimics the year 2012.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of:	**Measure Tested at Level of:**

(*must be consistent with levels entered in item S.26*)	
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☒ other: Bonnie Test Cases

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Not applicable.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

34 unique synthetic patient records were created in the Bonnie testing system for this measure. Cases were used to test the validity of each data element and timing relationship in the measure. Patient characteristics such as age, diagnosis, and medications were pre-determined to provide a variety of scenarios that adequately test patients passing each data element and failing each data element. Data included in cases and tested for this measure included diagnoses, medications, laboratory results, physical exams, patient orders, age, length of stay, and ED and inpatient encounters.

All 34 cases passed or failed as expected based on the data included in the case, confirming the measure logic is accurate and valid.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Not applicable. Bonnie test patients were developed for use across the different aspects of testing.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Not applicable.

2a2. RELIABILITY TESTING

Note: *If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*May be one or both levels*)

X **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

Not applicable. Synthetic test patients were created in the Bonnie testing environment.

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

The chart-abstracted version of this measure has been in national use since the 4th quarter of 2009. At the time the chart-abstracted measure was originally tested, extensive tests of measure reliability were conducted. At present, no performance data for the electronic version of the measure are currently available. Below is the reliability testing summary for NQF #0437 STK-04: Thrombolytic Therapy, from which this measure is derived.

At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with monthly data transmission to The Joint Commission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of approximately 900 patient records identified at 30 pilot sites over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rates of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify of potential measure enhancements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures.) The following is a list of the major tests conducted on the submitted data for this measure:

- Transmission of complete data
- Usage of data received (to understand if the healthcare organization provides the relevant service to treat the relevant population)
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity
- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors is as follows:

- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data element agreement rates reported to The Joint Commission for the time period of one year (4Q2010 – 3Q2011) have shown an overall agreement rate of 98.1%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-4.

Data Elements	total 'n' numerator	total 'n' denominator	agreement rate
Clinical Trial	712	714	99.7%
Date Last Known Well	181	206	87.9%
ED Patient	693	695	99.7%
Elective Carotid Intervention	711	714	99.6%
IV Thrombolytic Initiation	215	237	90.7%
IV Thrombolytic Initiation Date	25	27	92.6%
IV Thrombolytic Initiation Time	23	27	85.2%
Last Known Well	417	478	87.2%
Reason for Not Initiating IV Thrombolytic Therapy	155	202	76.7%
Time Last Known Well	157	198	80.0%

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Agreement rates for individual data elements tested for the chart-abstracted version of this measure were within acceptable levels. Once data are available for analysis, it is expected that reliability tests of the eCQM version of this measure will yield similar results.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

- ☒ **Critical data elements** (*data element validity must address ALL critical data elements*)
- ☒ **Performance measure score**
  - ☐ **Empirical validity testing**
  - ☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The Bonnie testing tool and environment were used to test the measure logic and value sets. Each data element and logic statement was tested to confirm actual results met expectations. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure. An example of negative testing would be to include test cases with pediatric ages to ensure that pediatric patients are not included in the measure.

Denominator test cases positively test to ensure patients with principal diagnosis of ischemic stroke admitted through the emergency department (ED), whose arrival time is within 2 hours of the 1) time they were known to

be at their baseline state of health or 2) time of symptom onset pass and fall into the measure. We have created negative test cases, with one example testing to ensure patients without a principal diagnosis of ischemic stroke do not actually make it into the denominator. Another example of negative testing is a test case that does not have an ED visit

Numerator test cases positively test to ensure patients receiving thrombolytic (t-PA) therapy within 3 hours of when it was witnessed or reported that the patient was last known to be without the signs and symptoms of the current stroke or at his or her baseline state of health fall into the measure. Negative test cases ensure that a patient who does not receive t-PA therapy in the specified time frame do not make it into the numerator population.

Denominator exception test cases for this measure ensure that patients are properly removed from the denominator for the following reasons:

- comfort measures
- intra-venous (IV) or intra-arterial t-PA therapy prior to arrival
- documentation of first National Institutes for Health Stroke Scale (NIHSS) with result of 0 in the ED
- medical reasons for not receiving IV t-PA
- prothrombin time > 15 seconds
- platelet count < 100,000
- INR > 1.7
- partial thromboplastin time > 40 seconds
- systolic blood pressure > 185 mmHg
- diastolic blood pressure > 110 mmHg
- patient refusal

Negative test cases are also run, an example of this would be comfort measures that are ordered but not in the specified time frame. Another example is negatively testing a patient with a prothrombin time < 15 seconds who does not qualify for a denominator exclusion. Testing confirmed patients meeting the exception criteria are removed from the measure appropriately.

A review of the measure specifications was also conducted to confirm the logic was properly expressed within the current version of the QDM and confirmed the logic matches the clinical intent of the measure, as stated in the measure header. This is done through use of a logic checklist to facilitate review of the measure logic, according to several checks, a few examples of which are included below:

- Is the intent of the measure described in the measure description articulated/ captured in the measure logic?
- Do the logic elements map to definitions in the measure narrative, data dictionary or supporting reference documentation?
- Do the populations in the narrative align with the populations defined in the logic?
- Are the mathematic inequalities reflective of the measure intent and represent the intended populations (for example: when intended the inequality represents less than rather than less and or equal to)?

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Bonnie results provide coverage and passing rates. This measure reached 100% coverage, confirming there is a test case for each pathway of logic. This measure also has a 100% passing rate, confirming all test cases performed as expected.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

This measure performed as expected in Bonnie.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Exclusions in the eCQM align with the chart-based version of the measure, and are clinically necessary for the interpretation of the measure. As noted previously, the chart-abstracted version of this measure has been in national use since the 4th quarter of 2009, and no data are available for the eCQM version of the measure. The below analysis addresses exclusions testing performed for the chart-abstracted version of the measure from which this measure is derived.

The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patient who have a Length of Stay greater than 120 days
- Patient enrolled in a Clinical Trial
- Patients admitted for Elective Carotid Intervention
- Time Last Known Well to arrival in the emergency department greater than 2 hours
- Patients with a documented Reason For Extending the Initiation of IV Thrombolytic
- Patients with a documented Reason For Not Initiating IV Thrombolytic

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are given below.

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%
10th Percentile: 0.24%
Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for Elective Carotid Intervention
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.5%
90th Percentile: 25.3%
Maximum: 95.2%

Exclusion: Time Last Known Well to arrival in the emergency department greater than 2 hours
None

Exclusion: Patients with a documented Reason For Not Initiating IV Thrombolytic
Overall Occurrence n = 73,966
Overall Occurrence Percentage: 24%
Minimum: 0.42%
10th Percentile: 5.5%
Median: 14.6%
90th Percentile: 54%
Maximum: 92%

Exclusion: Patients with a documented Reason For Extending the Initiation of IV Thrombolytic
Overall Occurrence = 2,208
Overall Occurrence Percentage: 0.70%
Minimum: 0.09%
10th Percentile: 0.25%
Median: 0.69%
90th Percentile: 2.3%
Maximum: 35.2%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Analysis of these data for the chart-abstracted progenitor of this measure indicated that all exclusions were appropriate. It is believed that results for exclusions in the eCQM will be similar when sufficient data have been received to perform such an analysis.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

- ☒ **No risk adjustment or stratification**
- ☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
- ☐ **Stratification by** Click here to enter number of categories **risk categories**
- ☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of an HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

A similar methodology will be used for the eCQMs.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

NQF#0437: STK-4 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10th Percentile
3Q2011	104	0.64221	0.40419	1	1	0.8	0.25	0
2Q2011	114	0.65061	0.40260	1	1	0.85714	0.33333	0
Q12011	116	0.6109	0.42108	1	1	0.775	0.0625	0
4Q2010	105	0.65572	0.40568	1	1	0.84615	0.33333	0
3Q2010	104	0.60829	0.42134	1	1	0.75	0	0
2Q2010	101	0.61933	0.38988	1	1	0.66667	0.33333	0
1Q2010	76	0.53939	0.41284	1	1	0.66667	0	0
4Q2009	39	0.54417	0.41298	1	1	0.625	0.03704	0

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

It should be noted that since data collection on this measure is completely voluntary for both The Joint Commission, the hospitals reporting on this measure are self-selected, and therefore, presumably have a particular interest and concern for improving stroke care. For this reason, the measure results demonstrated by this group most likely significantly overstates the rate for the population of all health care organizations.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Multiple data sources are not used for this measure.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Data not present in the structured field from which the measure draws will not be included in the measure calculation. In the Bonnie testing environment, missing data are tested as an expected "Fail" for that data element, and actual performance (whether or not the case fails) is compared to expectations to ensure missing data impacts data element and overall measure calculation as expected. This is the extent of missing data analysis that can be performed in Bonnie.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

See response for question 2b7.1 above.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

The measure performed as expected in the Bonnie testing environment.

3. Feasibility
Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
3a. Byproduct of Care Processes For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order). 3a.1. **Data Elements Generated as Byproduct of Care Processes.** Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score) If other:
3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment **Attachment:** STK4_eCQM_NQF_Measure_Feasibility_Assessment_Report.docx

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Not applicable.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
Value sets are housed in the Value Set Authority Center (VSAC), which is provided by the National Library of Medicine (NLM), in coordination with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

Viewing or downloading value sets requires a free Unified Medical Language System® (UMLS) Metathesaurus License, due to usage restrictions on some of the codes included in the value sets. Individuals interested in accessing value set content can request a UMLS license at (https://uts.nlm.nih.gov/license.html)

There are no other fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program http://jointcommission.org EHR Incentive Program https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

- Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
- Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 Medicare-certified hospitals (2015)

- Name of program and sponsor: Hospital Accreditation Program; The Joint Commission
- Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

- Name of program and sponsor: EHR Incentive Program; Centers for Medicare & Medicaid Services
- Purpose: The American Recovery and Reinvestment Act of 2009 (ARRA) (Pub.L. 111–5) was enacted on February 17, 2009. Title IV of Division B of ARRA amends Titles XVIII and XIX of the Social Security Act (the Act) by establishing incentive payments to eligible professionals (EPs), eligible hospitals, and critical access hospitals (CAHs), and Medicare Advantage Organizations to promote the adoption and meaningful use of interoperable health information technology (HIT) and qualified electronic health records (EHRs). These incentive payments are part of a broader effort under the HITECH Act to accelerate the adoption of HIT and utilization of qualified EHRs.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4,847 active registrations (September, 2015)

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not applicable.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance

results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

While the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are currently available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs, one of which is this measure. Data will be accepted through February 28th, 2017.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Not applicable.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0164 : Fibrinolytic Therapy received within 30 minutes of hospital arrival
0242 : Stroke and Stroke Rehabilitation: Tissue Plasminogen Activator (t-PA) Considered
0288 : Fibrinolytic Therapy Received Within 30 Minutes of ED Arrival
0437 : STK 04: Thrombolytic Therapy
1952 : Time to Intravenous Thrombolytic Therapy

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0242 : Stroke and Stroke Rehabilitation: Tissue Plasminogen Activator (t-PA) Considered
American Medical Association - Physician Consortium for Performance Improvement (AMA-PCPI) – no longer NQF endorsed

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Measures 0288 and 0164 are AMI (Acute Myocardial Infarction) measures. They are part of the Centers for Medicare & Medicaid Services/The Joint Commission aligned measures relating to the administration of fibrinolytic therapy for hospital inpatients and are harmonized with 0437 to the extent that the measures utilize some of the same data elements. The target population for 0288 and 0164 is inpatients with an ICD-10-CM Principal Diagnosis Code for acute myocardial infarction. The target population for 0437 differs in that it includes patients hospitalized for acute ischemic stroke. In addition, the evidence around the timeframe for administration of therapy is different for the AMI and ischemic stroke populations, and 0288 and 0164 include administration of lytic drugs other than activase/alteplase/IV t-PA/recombinant tissue plasminogen activator (rt-PA). Measure 0164 will be removed from the CMS/The Joint Commission aligned measures starting with 01/01/2016 discharges. The target population for measure 1952 from the American Heart Association/American Stroke Association also includes patients hospitalized for acute ischemic stroke; however, the measure captures average door-to-needle time and uses a target of less than 60 minutes rather than the proportion of patients who arrive within 2 hours and receive t-PA within 3 hours of time last known well. Measure 0242 is a physician performance measure with a targeted population of ischemic stroke patients identified through CPT codes and could extend to the outpatient setting. This measure evaluates physician practice as opposed to hospital processes. It is no longer NQF-endorsed. NQF#0437: STK 04: Thrombolytic Therapy: The measures are completely harmonized to the extent possible, given the fact that the data source for #0437 is the paper medical record, and the data source for #2834 is the electronic health record.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not applicable.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Ann, Watt, awatt@jointcommission.org, 630-792-5944-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Lisa, Anderson, landerson2@jointcommission.org, 630-792-5008-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel (TAP) is to provide advisory oversight in literature review, measure content and maintenance of the specifications.
Members are:

Harold P. Adams, Jr., MD
University of Iowa Health Care
Iowa City, IA

Mark J. Alberts, MD
University of Texas Southwestern
Dallas, TX

Anne W. Alexandrov, RN
Health Outcomes Institute
Fountain Hills, AZ

Nadine Allyn, RD
American Heart Association
Dallas, Texas

Mary G. George, MD
Centers for Disease Control and Prevention
Atlanta, GA

Martin Gizzi, MD
Legacy Health System
Portland, OR

Marianna Gorbaty
Mathematica Policy Research, Inc.
Washington, DC

Judith Hinchey, MD
Tuft's Medical School
Steward St. Elizabeth's Medical Center
Boston, MA

Edward C. Jauch, MD
Medical University of South Carolina
Charleston, SC 2946

Irene Katzan, MD
Cleveland Clinic
Cleveland, OH

Kimberly E. Levasseur-Franklin, PharmD
Northwestern Memorial Hospital
Chicago, IL

Kathy Morrison, RN
Penn State Hershey Medical Center
Hershey, PA

Marilyn M. Rymer, MD
University of Kansas
Kansas City, MO

Jeffrey Saver, MD
UCLA Medical Center
Los Angeles, CA

Lee H. Schwamm, MD
Harvard Medical School
Massachusetts General Hospital
Boston, MA

Deborah, Summers, RN
St. Luke's Brain and Stroke Institute
Kansas City, MO

Osama O. Zaidat, MD
St. Vincent Mercy Hospital
Toledo, OH

Richard D. Zorowitz, MD
Medstar National Rehabilitation Network
Washington, DC

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2012

Ad.3 **Month and Year of most recent revision:** 04, 2015

Ad.4 **What is your frequency for review/update of this measure?** Annual

Ad.5 **When is the next scheduled review/update for this measure?** 04, 2016

Ad.7 **Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. The measures and specifications are provided without warranty.

Ad.8 **Additional Information/Comments:** Not applicable.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2835
Measure Title: STK 05: Antithrombotic Therapy By End of Hospital Day Two
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of ischemic stroke patients who had antithrombotic therapy administered by end of hospital day two (with the day of arrival being day 1).

This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-4: Thrombolytic Therapy, STK-6: Discharged on Statin Medication, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs. STK-5, Antithrombotic Therapy By End of Hospital Day Two, is one of six of the measures in this set that have been reengineered as eCQMs and are included in the EHR Incentive Program and Hospital Inpatient Quality Reporting Program.
Developer Rationale: Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. The effectiveness of early antithrombotic administration in reducing stroke mortality recurrence rates has been demonstrated in several large clinical trials. Data suggest that antithrombotic therapy should be administered by the end of hospital day two to reduce stroke mortality and morbidity as long as no contraindications exist.

Healthcare organizations that track antithrombotic therapy prescribed by the end of day two for internal quality improvement purposes have seen an improvement in the measure rate over time. The chart-abstracted measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

Numerator Statement: Patients who had antithrombotic therapy administered the day of or day after hospital arrival.
Denominator Statement: Patients with a principal diagnosis of Ischemic stroke.
Denominator Exclusions: Denominator Exclusions:

- Patients who have a duration of stay less than 2 days
- Patients with comfort measures documented on day or the day after arrival
- Patients with intra-venous or intra-arterial Thrombolytic (t-PA) Therapy administered within 24 hours prior to arrival or anytime during hospitalization.

Denominator Exceptions:

- Patients with a documented reason for not administering antithrombotic therapy the day of or day after hospital arrival.

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Pharmacy
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary:

Evidence for this new eMeasure is the same as the existing measure #0438 STK 05: Antithrombotic Therapy By End of Hospital Day Two:

- The body of evidence consistently supports that early antithrombotic therapy reduces the risk of non-fatal stroke and death for patients who have experienced an acute ischemic stroke. There is no evidence refuting the benefit of this recommendation.
- Class I, Level of Evidence A – AHA ASA 2007 Guidelines for the Early Management of Adults with Ischemic Stroke, page 1681.
    - The oral administration of aspirin (initial dose is 325 mg) within 24 to 48 hours after stroke onset is recommended for treatment of most patients (Class I, Level of Evidence A). This recommendation has changed in that a dose of aspirin is now included.

Questions for the Committee:
The developer attests the underlying evidence for the measure #0438 (chart-abstracted version) has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the chart-abstracted measure has not changed and can be applied to the eMeasure; therefore there is no need for repeat discussion and voting on Evidence?

Guidance from the Evidence Algorithm: Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency: high (Box 5) → Box 5a High

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**
Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer states no performance data for the electronic version of this measure are yet available.
- This eCQM is currently included in the Hospital Inpatient Quality Reporting Program (IQR) and EHR Incentive program.
    - In CY2016, CMS is requiring organizations participating in HIQR to electronically submit one quarter of data on 4 of 28 available eCQMs. This measure is 1 of the 28 eCQMs.
- The developer also provided performance data from the chart abstracted measure which seems to be very high with little to no room for improvement.

Disparities:

- The developer does not provide disparities data from use of this measure.
- Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-

Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group; however, this risk decreases as people age."

Questions for the Committee:

- *Is there a gap in care that warrants a national performance eMeasure?*

  How can this measure be used to better understand disparities in care and outcomes for certain groups?

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **Similar to #0438 chart abstracted this is eMeasure.

Agree-level of evidence has not changed. No need repeat discussion from #0438.

**evidence supports eMeasure

**This eMeasure relies on the same body of evidence as previously-endorsed NQF measure 438. I am not aware of any new information that changes the evidence base.

**Evidence same as 0438. No change in evidence. No vote needed.

1b. Performance Gap

Comments: **no data for this eMeasure

**Performance data not available but chart abstracted performance data high participation

**No performance data for the eMeasure is available, but one would expect high performance with little, if any, room for improvement, based on the chart-abstracted measure.

Data could be analyzed/stratified by sociodemographic variables to identify disparities in care.

**No data presented for eMeasure. Disparities in stroke incidence and care are noted elsewhere. These are not directly addressed in the data for this measure. The presence and use of this measure and 0438 should enable collection of disparities data for future reviews.

1c. High Priority (previously referred to as High Impact)

Comments: **Quality construction easy to assess.

**NA

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic Clinical Data, Electronic Health Record (EHR), Imaging/Diagnostic Study, Laboratory Data, and Pharmacy Data. This is an eMeasure.

Specifications:

- HQMF specifications for eMeasure are included in the document set on SharePoint. See eMeasure Technical Review below.
- The developer states that changes have been made to the eCQM specifications to reflect revisions to the chart abstracted measure from which this measure is derived.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*

- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor Review:

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously Demonstrated through Bonnie testing
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible. This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment.

2a2. Reliability Testing **Testing attachment**

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☒ **No**

Method(s) of reliability testing:

- Dataset used for testing included 31 synthetic records created in the Bonnie testing system. Measure not tested in EHR.
- Data element validity testing was performed and will count for data element reliability as well – see validity testing section.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results from the Bonnie tool demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm: Precise specifications (Box 1) →empiric reliability testing as specified (Box 2) → empiric validity testing of patient-level data (Box 3) → YES: Use rating from validity testing of patient-level data elements (Box 10)→Assessed the measure logic only →Low certainty or confidence data used in measure are valid (Box 12b)→Low (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Preliminary rating for reliability: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**
☒ **Empirical validity testing of the measure score**

Validity testing method:

- The Bonnie testing tool and environment with 31 synthetic patient records were used to test the measure logic and value sets. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure.
- The numerator, denominator, exclusions/exceptions, and logic statement were tested to confirm actual results met expectations.
- The developer also conducted a review of the measure specifications to confirm the measure logic was properly expressed with the current version of the QDM and that the logic matches the clinical intent of the measure.

Validity testing results:

- The testing results from the Bonnie tool reached 100% coverage and confirmed there was a test case for each pathway of logic (negative and positive test cases).
- The measure also had a 100% passing rate which confirmed that all the test cases performed as expected.
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer submitted additional information on exclusions and meaningful differences.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Is it reasonable to assume that the impact of exclusions will be similar for the eMeasure and the chart-abstracted measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer submitted additional information on exclusions and meaningful differences.

Question for the Committee:

- *For meaningful differences in the chart-abstracted measure the developer refers to the performance scores over time. Does the SC think the eMeasure will demonstrate similar results to the chart-abstracted measure which show little to no room for improvement?*

2b6. Comparability of data sources/methods: Not Applicable

2b7. Missing Data

- The developer describes the handling of missing data but does not provide information on the frequency of missing data or potential impact on measure results due to limited testing abilities using the Bonnie tool.

Guidance from Validity Algorithm: Specifications consistent with evidence (Box 1) → threats to validity empirically assessed to some extent (Box2) → empirical testing of data elements using BONNIE tool (Box 3,10) → Moderate

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **NA

**Acceptable

**Again all specifications seem appropriate and positive. As noted above, varying use of different antithrombolitics and doses may make validity appraisal more problematic.

2a2. Reliability Testing

Comments: **See above. Small sample size.

Is this adequate to generalize for widespread implementation? Unknown...

**Not concerned about not testing face validity

**No validity data from eMeasure use. 31 Bonnie test results use. All indications are very positive. Real data remain to be seen.

2b2. Validity Testing

Comments: **Agree with measure worksheet-insufficient data to rate validity.

**Numerous threats to validity

**Exclusions align with chart data. No risk adjustments. Aspirin effects are identified as modest but positive. There is a plan for dealing with missing data, but no frequency data are presented.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Reliability based on small synthetic data set limits reliability assessment.

**Low

**Reliability testing appears to be insufficient to truly judge the performance of the eMeasure on actual EHR data.

**There are no direct data from eMeasure use, but data from 0438 seem appropriate. Bonnie test results on 31 cases are positive - though this number is relatively small. Test with real (vs. synthetic) cases is preferable.

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is EHR; all data elements are in defined fields in EHR.
- The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible.
- This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment.
- Per developer, they were unable to directly assess data accuracy of the data elements in an EHR system. Workflow or the degree to which the data elements are captured during the course of care was also not assessed.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Clinical data should be routinely generated-unknown how accurate abstraction might be.

**Acceptable

**I am concerned that exclusions will not be identified correctly without manual abstraction.

I am not certain all data elements are defined fields in EHR

**This is a legacy measure. EHR data should be readily available. Data accuracy are not assessed. Are exclusions in the EHR accurate.

There is data entry time and data extraction/review time. Is there an appropriate cost/benefit ratio to the use of this measure?

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- CMS Hospital Inpatient Quality Reporting Program (IQR)/EHR Incentive Program: The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. The EHR Incentive Program provides incentive payments to promote the adoption and meaningful use of interoperable HIT and qualified EHRs.

- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results:

- The developer states that while the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are currently available.
- In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs, one of which is this measure. Data will be accepted through February 28th, 2017.

Unexpected findings (positive or negative) during implementation:

- Developer did not identify unexpected findings during implementation.

Potential harms:

- Developer did not identify any unintended consequences related to this measure.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Not currently publicly reported. Paper measure is in use. in use.

**No performance data but chart abstracted usability acceptable

**eMeasure is not being publicly reported but chart-abstracted measure is.

**This measure is a part of accountability programs. It is assumed that results in more frequent administration of anti-thrombolitics before end of day two. That should result in improved care. There is no evidence of unintended consequences (assuming appropriate identification of exclusions) other than the time and effort to record and review data.

Criterion 5: Related and Competing Measures

Related or competing measures:

- 0438 : STK 05: Antithrombotic Therapy By End of Hospital Day Two
- 0068 : Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet
- 0435 : STK 02: Discharged on Antithrombotic Therapy
- 0438 : STK 05: Antithrombotic Therapy By End of Hospital Day Two

- 0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy – American Academy of Neurology – not NQF endorsed

Harmonization:

- Measure 0435, Discharged on Antithrombotic Therapy, is the second (STK-2) measure in The Joint Commission stroke core measure set and also targets the ischemic stroke population; however, the timeframe for antithrombotic administration is different in the two measures. #0435 focuses on the prescription of antithrombotic medications at the time of hospital discharge. All other data elements are aligned between

the two measures.

- Measure 0068 is a physician level measure with a different target population - patients with ischemic vascular disease who were discharged alive for acute myocardial infarction (AMI), coronary artery bypass graft (CABG) or percutaneous coronary interventions (PCI).
- NQF#0438: STK 05: Antithrombotic Therapy By End of Hospital Day Two: The measures are completely harmonized to the extent possible, given the fact that the data source for #0438 is the paper medical record, and the data source for #2835 is the electronic health record.

Pre-meeting public and member comments

- No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0438 **NQF Project:** Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. **(evaluation criteria)**

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)***:**
The focus of the measure is to decrease stroke recurrence rates for ischemic stroke patients by administering aspirin or another antithrombotic medication within 2 days of symptom onset.
Antithrombotic administered by end of hospital day 2 >> secondary stroke prevention >> decreased morbidity and mortality.

1c.2-3 Type of Evidence *(Check all that apply)***:**
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)***:**
This measure is consistent with the guidelines recommended by the American Heart Association/American Stroke Association for the early management of adults with ischemic stroke. A small but statistically significant decline in risk of mortality and morbidity has been demonstrated when aspirin is administered within 48 hours after onset of stroke. The primary effects of aspirin are due to reduction of early recurrent stroke rather than limitation of neurological consequences of the stroke. The focus of both the performance measure and the body of evidence supports the need for early administration of antithrombotic therapy.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: The two most frequently cited trials in the literature are the International Stroke Trial (i.e., 300 mg of aspirin daily versus control; 1994) and the Chinese Acute Stroke Trial(i.e., 160 mg aspirin daily versus control; 1997). Prior to these trials, there was little to no information about the effects of antiplatelet agents on acute ischemic stroke.

In 2002, the Antithrombotic Trialists' Collaboration published a systematic overview of the relevant literature supporting antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. The collaboration reviewed 287 studies involving 135,000 patients in comparisons of antiplatelet therapy versus control and 77,000 patients in comparisons of different antiplatelet regimens. Relevant trials were identified through electronic database searches (MEDLINE, Embase, Derwent, Scisearch, and Biosis), searching the trial registers of the Cochrane Stroke and Peripheral Vascular Disease Groups; manual searching of journals, abstracts, and proceedings of meetings; scrutinizing the reference lists of trials and review articles; and professional inquiry, including colleagues and representatives of pharmaceutical companies.

The meta-analysis identified seven randomized control trials encompassing 40,821 patients that focused on the effects of antithrombotic therapy in acute ischemic stroke patients. The patients enrolled in these seven trials received on average three weeks of antiplatelet therapy following the onset of acute stroke symptoms. Overall, antiplatelet therapy resulted in a proportional reduction of 11% in vascular events (i.e., non-fatal stroke or death).

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting early antithrombotic administration for patients with acute ischemic stroke is high. Multiple large randomized control studies have been published that studied the effects of aspirin, aspirin in combination with other antithrombotic agents, and alternative antithrombotic agents as monotherapy. As noted by the AHA/ASA and the American Academy of Neurology, the early administration of antithrombotic therapy, the focus of this measure, has been found to protect against stroke recurrence in the initial weeks following ischemic stroke. The administration of aspirin is not a substitute for other acute interventions, especially intravenous administration of thrombolytic therapy (IV t-PA), for the treatment of acute ischemic stroke. The administration of aspirin as an adjunctive therapy within 24 hours of IV t-PA is also not recommended.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of evidence consistently supports that early antithrombotic therapy reduces the risk of non-fatal stroke and death for patients who have experienced an acute ischemic stroke. There is no evidence refuting the benefit of this recommendation.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
According to the Antithrombotic Trialists' Collaboration meta-analysis, there is strong evidence on the immediate hazards and net benefits of early antithrombotic administration for the prevention of stroke recurrence. Four of the seven trials included in the trialists' 2002 meta-analysis separated stroke outcomes into those considered to be due to hemorrhagic events and those of ischemic or unknown origin. Antithrombotic therapy produced an absolute excess of 1.9 hemorrhagic strokes per 1,000 patients, which was counterbalanced by an absolute reduction of 6.9 fewer ischemic strokes per 1,000, yielding an overall risk reduction in the risk of any further stroke of 5.4 per 1,000 patients. The excess risk of major extracranial hemorrhage was estimated at about three excess bleeds per 1,000 patients receiving early antithrombotic therapy. An increased risk of bleeding was associated with concurrent heparin administration (i.e., absolute excess 9 bleeds per 1,000 with heparin versus 2 bleeds per 1,000 without heparin.

As previously stated, the benefits of early antithrombotic therapy outweigh the risks of hemorrhage. Aspirin is a cost-effective therapy (Gaspoz, et al., 2002).

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: American Heart Association. During our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, and provided information that was accessible and met the requirements set out in this measure maintenance form.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.13 Grade Assigned to the Body of Evidence: Class I, Level of Evidence A

1c.14 Summary of Controversy/Contradictory Evidence: The benefit of early antithrombotic therapy post ischemic stroke is undisputed. No position against early antithrombotic therapy was noted in the literature; however, the substitution of early antithrombotic therapy for other acute interventions, (i.e., IV t-PA) is not recommended. Additionally, the administration of antithrombotic medications in conjunction with other antithrombotic agents (e.g., heparin) or soon after IV t-PA administration is inadvisable.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and thrombolytic therapy for ischemic stroke: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest. 2004; 126;483S-512S.
- Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomized trials of anti-platelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients [published correction appears in BMJ. 2002;324:141]. BMJ. 2002;324:71– 86.
- Brott TG, Clark WM, Grotta JC, et al. Stroke the first hours. Guidelines for acute treatment. Consensus Statement. National Stroke Association. 2000.
- Chen ZM, Sandercock P, on behalf of the AntiThrombotic Trialists Collaboration (ATT). Indications for early aspirin use in acute stroke: a systematic overview of over 40,000 randomised patients. Cerebrovasc Dis. 1998;8(suppl 4):38.
- Chen ZM, Sandercock P, Pan HC, et al. Indications for early aspirin use in acute ischemic stroke: a combined analysis of 40,000 randomized patients from the Chinese acute stroke trial and the international stroke trial. On behalf of the CAST and IST collaborative groups, Stroke 2000;31:1240-1249.
- Coull BM, Williams LS, Goldstein LB, et al. Anticoagulants and Antiplatelet Agents in Acute Ischemic Stroke. Report of the Joint Stroke Guideline Development Committee of the American Academy of Neurology and the American Stroke Association (a Division of the American Heart Association) Stroke. 2002;33:1934 -1942.
- Furie KL, Kasner SE, Adams RJ, Albers GW, et al. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2011;42:246-49.
- Gaspoz J, Coxson PG, Goldman PA, Williams LW, Kuntz KM, Hunink MGM, Goldman L. Cost effectiveness of aspirin, clopidogrel, or both for secondary prevention of coronary heart disease. NEJM. 2002;346:1800-1806.
- Guyatt GH, Akl EA, Crowther M, Gutterman DD, Sch?nemann HJ, and for the American College of Chest Physicians Antithrombotic Therapy and Prevention of Thrombosis Panel. Executive summary: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141: 34S.
- Johnson ES, Lanes SF, Wentworth CE, Satterfield MH, Abebe BL, Dicker LW. A meta-regression analysis of the dose-response effect of aspirin on stroke. Arch Intern Med. 1999;159:1248 –1253.
- Sandercock P, Collins R, Counsell C, Farrell B, Peto R, Slattery J, Warlow C, International Stroke Trial Collaborative Group: The International Stroke Trial (IST): A randomised trial of aspirin, subcutaneous heparin, both, or neither among 19,435 patients with acute ischaemic stroke. Lancet. 1997;349:1569-1581.
- The ESPS Group. The European Stroke Prevention Study (ESPS): principal end-points. Lancet. 1987:2:1351–1354.

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)*:
AHA/ASA 2007 Guidelines for the Early Management of Adults with Ischemic Stroke, page 1681.
Class I, Level of Evidence A
1. The oral administration of aspirin (initial dose is 325 mg) within 24 to 48 hours after stroke onset is recommended for treatment of most patients (Class I, Level of Evidence A). This recommendation has changed in that a dose of aspirin is now included.

1c.17 Clinical Practice Guideline Citation: Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A Guideline From the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1681.

1c.18 National Guideline Clearinghouse or other URL:
http://www.guidelines.gov/search/search.aspx?term=early+management+of+adults+with+ischemic+stroke

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: This guideline was approved by the American Heart Association Science Advisory and Coordinating Committee on January 6, 2007. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.23 **Grade Assigned to the Recommendation:** Class I, Level of Evidence A

1c.24 **Rationale for Using this Guideline Over Others:** The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: High 1c.26 Quality: High1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

0438_Evidence_MSF5.0_Data-635827722640851821.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*

Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. The effectiveness of early antithrombotic administration in reducing stroke mortality recurrence rates has been demonstrated in several large clinical trials. Data suggest that antithrombotic therapy should be administered by the end of hospital day two to reduce stroke mortality and morbidity as long as no contraindications exist.

Healthcare organizations that track antithrombotic therapy prescribed by the end of day two for internal quality improvement purposes have seen an improvement in the measure rate over time. The chart-abstracted measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This measure is a legacy eCQM that is currently included in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data for the electronic version of the measure are yet available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs. This measure is one of the 28.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
In October, 2009, The Joint Commission added the chart-abstracted stroke (STK) measure set as a new core measure option to meet performance measure requirements for Joint Commission hospital accreditation purposes. Below is the specified level of analysis for STK-5 beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 1955 denominator cases; 1875 numerator cases; 49 hospitals; 0.95908 national aggregate rate; 0.95947 mean of hospital rates; 0.10574 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.94949 50th percentile rate/median rate; 0.85165 25th percentile rate/lower quartile; and, 0.66667 10th percentile rate.

CY 2010: 18,728 denominator cases; 18,223 numerator cases; 137 hospitals; 0.97304 national aggregate rate; 0.95924 mean of hospital rates; 0.08189 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.98361 50th percentile rate/median rate; 0.96183 25th percentile rate/lower quartile; and, 0.90909 10th percentile rate.

CY 2011: 22,328 denominator cases; 21,814 numerator cases; 157 hospitals; 0.97698 national aggregate rate; 0.95933 mean of hospital rates; 0.10243 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.98667 50th percentile rate/median rate; 0.96053 25th percentile rate/lower quartile; and, 0.928 10th percentile rate.

CY 2012: 22,258 denominator cases; 21,866 numerator cases; 158 hospitals; 0.98239 national aggregate rate; 0.98042 mean of hospital rates; 0.03174 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99396 50th percentile rate/median rate; 0.97143 25th percentile rate/lower quartile; and, 0.95098 10th percentile rate.

CY 2013: 33,651 denominator cases; 33,133 numerator cases; 263 hospitals; 0.98461 national aggregate rate; 0.97976 mean of hospital rates; 0.04874 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99422 50th percentile rate/median rate; 0.9781 25th percentile rate/lower quartile; and, 0.95745 10th percentile rate.

CY 2014: 162,275 denominator cases; 159.848 numerator cases; 1300 hospitals; 0.98504 national aggregate rate; 0.97832 mean of hospital rates; 0.07287 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99261 50th percentile rate/median rate; 0.97832 25th percentile rate/lower quartile; and, 0.95707 10th percentile rate.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers,

immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al., 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

In the national REGARDS cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years of age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age : RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer, et al., 2010).

In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).
In terms of early antithrombotic administration, no significant disparities were noted across studies. Recent cross-sectional studies from nationwide health surveys suggest that once patients have a stroke, the use of most risk-reduction strategies does not vary significantly by race. An analysis of 1045 veteran patients found the use of aspirin similar for blacks or African Americans and whites, although the intervention was underutilized with aspirin administration in approximately 74% for both groups (Ross and Bravata, 2009). Similarly, data from the Centers for Disease Control (CDC) Behavioral Risk Factor Surveillance System (BRFSS) showed that black or African American stroke survivors were no less likely to be prescribed aspirin than whites.

Since the last endorsement date, several studies have reported results similar to those from BRFSS in 2009. Schwamm and colleagues (2010) found significant and important differences in quality of care related to early antithrombotics when multivariate models were constructed adjusting for patient-level characteristics only. Race/ethnicity Black versus White: unadjusted OR 0.96 [95% CI 0.92-

1.00]; adjusted for patient characteristics OR 0.90 [95% CI 0.86-0.94]; adjusted for patient and hospital characteristics OR 0.97 [95% CI 0.91-1.02]. Race/ethnicity Hispanic versus White: unadjusted OR 0.95 [95% CI 0.88-1.02]; adjusted for patient characteristics OR 0.89 [95% CI 0.83-0.96]; adjusted for patient and hospital characteristics OR 0.96 [95% CI 0.88-1.05]. Total N=271,769; All n 94.23%; White 94.28%; Black 94.08%; Hispanic 93.98%.

A more recently published study (Qian F, et al, 2013) from GWTG found that non-Hispanic black patients were less likely to receive early antithrombotics when compared to other race/ethnicity groups. Using patient data (n=200,900) from the American Heart Association/American Stroke Association Get With The Guidelines (GWTG)-Stroke program from April 2003 through December 2008, the following performance measure rates were reported for early antithrombotic therapy: non-Hispanic White (n=170,694) 94.9%; non-Hispanic Black (n=20,514) 94.0%; Hispanic (n=6632) 94.3%; and non-Hispanic Asian American (n=3060) 94.8% Data from the Paul Coverdell National Acute Stroke Registry (PCNASR) for early antithrombotic therapy were similar (n=58,823): White 96.5%; Other Race 95.6% (Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, 2014).

• Ambriz EH, Woodard LD, Kressin NR, Petersen LA. Use of smoking cessation interventions and aspirin for secondary prevention: are there racial disparities? Am J Med Qual. 2004; 19:166-171.
• Bhattacharya P, Mada F, Salowich-Palm L, Hinton S, Millis S, Watson SR, Chaturvedi S, Rajamani K. Department of neurology and Stroke Program, Wayne State University School of Medicine, Detroit, Michigan. Are racial disparities in stroke care still prevalent in certified stroke centers? J Stroke Cerebrovasc Dis. 2011.
• Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, Yan T, Escarce JJ. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-8.
• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, Griffith P, Gorelick PB, Howard G, Leira EC, Morgenstern LB, Ovbiagele B, Peterson E, Rosamond W, Trimble B, Valderrama AL, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• Gillum RF,Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 200;42(11):3294-6.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-75.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Jacobs BS, Birbeck G, Mullard AJ, Hickenbottom S, Kothari R, Roberts S, Reeves MJ. Quality of hospital care in African Americans and white patients with ischemic stroke and TIA. Neurology.2006;66:809-14.
• Karve S, Balkrishnan R, Seiber E Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, Kerber KA, Skolarus LE, Meurer WJ, Burke JF, Adelman EE, Baek J, Lisabeth LD. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol. 2013;74:778–785.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Ross JS, Bravata DM. Use of stroke secondary prevention services: are there disparities in care? Stroke. 2009;40:1811-1819.
• Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A,

Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Schwamm LH, Syed FA, Reeves MJ, Smith EE, Saver JL, Messe S, Bhatt DL, Grau-Sepulveda MV, Peterson ED, Fonarow GC. Temporal trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With the Guidelines-Stroke hospitals. Circ Cardiovasc Quality Outcomes. 2013;6:543-549.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States remains attributable to stroke. More women than men die of stroke each year. Women accounted for almost 60% of US stroke deaths (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, et al., 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 (P<0.05) (US Burden of Disease Collaborators, 2013) . Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

Antithrombotic agents significantly reduce the incidence of a recurrent vascular event after a stroke. Among these agents, aspirin has been the drug most widely studied. The International Stroke Trial demonstrated that antithrombotic administration, specifically aspirin, within the first 48 hours after stroke, significantly reduced the risk recurrent ischemic stroke and death (11.3% vs. 12.4%) in the first 14 days following the event. In conjunction with this, findings from the Chinese Acute Stroke Trial indicate that aspirin produces a modest reduction of approximately 10 deaths per 1000 during the first few weeks. Both trials recommend that aspirin should be given as soon as possible after the onset of stroke symptoms. It appears that the primary benefits of aspirin are due to early reduction in recurrent stroke rather than limitation of neurological deficits of the first stroke. For early antithrombotic therapy, oral administration of aspirin (initial dose is 325 mg) within 24 to 48 hours after stroke onset remains the current clinical guideline recommendation for treatment of most patients (Jauch, E. C., et al., 2013; Sandercock, P. A., et. al., 2014).

1c.4. Citations for data demonstrating high priority provided in 1a.3

• Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A Guideline From the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1655-1711.
• Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and Thrombolytic Therapy for Ischemic Stroke. Chest Vol. 119, 2001: 300-320.
• Brott TG, Clark WM, Grotta JC, et al. Stroke the first hours. Guidelines for acute treatment. Consensus Statement. National Stroke Association. 2000.
• Chen ZM, Sandercock P, on behalf of the AntiThrombotic Trialists Collaboration (ATT). Indications for early aspirin use in acute stroke: a systematic overview of over 40,000 randomised patients. Cerebrovasc Dis. 1998;8(suppl 4):38.

• Chen ZM, Sandercock P, Pan HC, et al. Indications for early aspirin use in acute ischemic stroke: a combined analysis of 40,000 randomized patients from the Chinese acute stroke trial and the international stroke trial. On behalf of the CAST and IST collaborative groups, Stroke 2000;31:1240-1249.
• Coull BM, Williams LS, Goldstein LB, et al. Anticoagulants and Antiplatelet Agents in Acute Ischemic Stroke. Report of the Joint Stroke Guideline Development Committee of the American Academy of Neurology and the American Stroke Association (a Division of the American Heart Association) Stroke. 2002;33:1934 -1942.
• George, M., Xin, Tong, McGruder, H., Yoon, P., Rosamond, W., Winquist, A., Hinchey, J., Wall, H., Pandey, D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Guyatt GH, Akl EA, Crowther M, Gutterman DD, Schu?nemann HJ, and for the American College of Chest Physicians Antithrombotic Therapy and Preveniton of Thrombosis Panel. Chest. 2012;141: 33S.
• Guideline on the Use of Aspirin as Secondary Prophylaxis for Vascular Disease in Primary Care, Centre for Health Services Research University of Newcastle upon Tyne, & Centre for Health Economics of York, 1998.
• Jauch EC, Saver JL, Adams HP, Bruno A, Connors JJ, Demaerschalk BM, Khatri P, McMullan PW, Qureshi AI, Rosenfield K, Scott PA, summers DR, Wang DZ, Wintermark M, Yonas, H. Guidelines for the early management of patients with acute ischemic stroke: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44:893-899.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Sandercock P, Collins R, Counsell C, Farrell B, Peto R, Slattery J, Warlow C, International Stroke Trial Collaborative Group: The International Stroke Trial (IST): A randomised trial of aspirin, subcutaneous heparin, both, or neither among 19,435 patients with acute ischaemic stroke. Lancet. 1997;349:1569-1581.
• Sandercock P, Counsell C, Tseng MC, Cecconi E. Oral antiplatelet therapy for acute ischaemic stroke. The Cochrane Database of Systematic Reviews. 2014;3:CD000029.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/ecqm_library.html

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is an eMeasure **Attachment:** STK5_MAT.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** SKT5Stroke_v4_Mon_Apr_06_11.03.30_CDT_2015.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Changes have been made to the eCQM specifications in order to reflect revisions to the chart abstracted measure from which this measure is derived.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients who had antithrombotic therapy administered the day of or day after hospital arrival.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Antithrombotic Therapy
- Antithrombotic Therapy is represented with the QDM datatype and value set of Medication, Administered: Antithrombotic Therapy (OID: 2.16.840.1.113883.3.117.1.7.1.201)

Non-Elective Inpatient Encounter
- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

Emergency Department Visit
- Emergency Department Visit is represented with the QDM datatype and value set of Encounter, Performed: Emergency Department Visit (OID: 2.16.840.1.113883.3.117.1.7.1.292)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients with a principal diagnosis of Ischemic stroke.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Principal Diagnosis of Ischemic Stroke
- Ischemic Stroke is represented with the QDM datatype and value set of Diagnosis, Active: Ischemic Stroke (OID: 2.16.840.1.113883.3.117.1.7.1.247)

- Ordinality: Principal (OID: 2.16.840.1.113883.3.117.1.7.1.14)

Non-Elective Inpatient Encounter
- Non-Elective Inpatient Encounter is represented with the QDM data type and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Denominator Exclusions:
- Patients who have a duration of stay less than 2 days
- Patients with comfort measures documented on day or the day after arrival
- Patients with intra-venous or intra-arterial Thrombolytic (t-PA) Therapy administered within 24 hours prior to arrival or anytime during hospitalization.

Denominator Exceptions:
- Patients with a documented reason for not administering antithrombotic therapy the day of or day after hospital arrival.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Denominator Exclusions:
Non-Elective Inpatient Encounter
- Non-Elective Inpatient Encounter is represented with the QDM data type and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

Emergency Department Visit
- Emergency Department Visit is represented with the QDM datatype and value set of Encounter, Performed: Emergency Department Visit (OID: 2.16.840.1.113883.3.117.1.7.1.292)

Thrombolytic (t-PA) Therapy Already Administered
- Thrombolytic (t-PA) Therapy is represented with either of the following QDM datatypes and value sets:
  - Medication, Administered: Thrombolytic (t-PA) Therapy (OID: 2.16.840.1.113883.3.117.1.7.1.226)
  - Procedure, Performed: Intravenous or Intra-arterial Thrombolytic (t-PA) Therapy (OID: 2.16.840.1.113762.1.4.1045.21)

Comfort Measures
- Comfort Measures are represented with the QDM datatypes and value set of:
- Intervention, Order: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)
- Intervention, Performed: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)

Denominator Exceptions:
Non-Elective Inpatient Encounter
- Non-Elective Inpatient Encounter is represented with the QDM data type and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

Emergency Department Visit
- Emergency Department Visit is represented with the QDM datatype and value set of Encounter, Performed: Emergency Department Visit (OID: 2.16.840.1.113883.3.117.1.7.1.292)

Reasons for Not Administering Antithrombotic Therapy
- Antithrombotic Ingredient Specific Medication is represented with the QDM datatype and value set of Medication, Order: Antithrombotic ingredient specific (OID: 2.16.840.1.113762.1.4.1021.8)
- Antithrombotic Medication is represented with the QDM datatype and value set of Medication, Administered: Antithrombotic Therapy (OID: 2.16.840.1.113883.3.117.1.7.1.201)
- Medication, Order not done: Medical Reason (OID: 2.16.840.1.113883.3.117.1.7.1.473)
- Medication, Administered not done: Medical Reason (OID: 2.16.840.1.113883.3.117.1.7.1.473)

- Medication, Order not done: Patient Refusal (OID: 2.16.840.1.113883.3.117.1.7.1.93)
- Medication, Administered not done: Patient Refusal (OID: 2.16.840.1.113883.3.117.1.7.1.93)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable, the measure is not stratified.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
See attached HQMF file.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.

eMeasures are calculated using only the structured data collected in certified EHR technology (CEHRT). Data not present in the structured field from which the measure draws will not be included in the measure calculation.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Pharmacy

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Hospitals report EHR data using Certified Electronic Health Record Technology (CEHRT), and by submitting Quality Reporting Document Architecture Category 1 (QRDA-1).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF0438_CMS72v4_STK5_Bonnie_Testing.xlsx,STK5_eCQM_testing_attachment-635907890708491062.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2835

Measure Title: STK05: Antithrombotic Therapy By End of Hospital Day Two

Date of Submission: 1/14/2016

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	X Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on

testing to demonstrate meeting the sub criteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):
- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR
- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-

item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒other: Bonnie Test Cases	☒ other: Bonnie Test Cases (see question 1.2)

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

This measure is a legacy eCQM (an NQF endorsed measure that has been re-specified into an eMeasure and is currently used in a federal quality program/programs). In accordance with the modified NQF requirements for testing of eMeasures, the Bonnie testing tool was used to simulate a testing environment where measure specifications and HQMF output are tested against synthetic test data. Measure developers rely on the results in Bonnie to confirm whether the measure logic is performing as expected. Reference the eCQI Resource Center website (https://ecqi.healthit.gov/ecqm-tools/tool-library/bonnie) or the Bonnie testing tool website (https://bonnie.healthit.gov/) for more information about Bonnie functionality and its role in measure development. Please also reference the Bonnie testing worksheet attachment for detailed Bonnie test cases and testing results for this measure.

While Bonnie does consume HQMF specifications, we are not able to test the measure with data from HQMF implemented in EHRs.

1.3. What are the dates of the data used in testing? The measure specifications were tested in the Bonnie testing environment which mimics the year 2012.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☒ other: Bonnie Test Cases

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Not applicable.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

31 unique synthetic patient records were created in the BONNIE testing system for this measure. Cases were used to test the validity of each data element and timing relationship in the measure. Patient characteristics such as age, diagnosis, and medications were pre-determined to provide a variety of scenarios that adequately test patients passing each data element and failing each data element. Data included in cases and tested for this measure included diagnoses, medications, procedures, patient orders, age, length of stay, and ED and inpatient encounters.

All 31 cases passed or failed as expected based on the data included in the case, confirming the measure logic is accurate and valid.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Not applicable. Bonnie test patients were developed for use across the different aspects of testing.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Not applicable.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*May be one or both levels*)

X **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

The chart-abstracted version of this measure has been in national use since the 4th quarter of 2009. At the time the chart-abstracted measure was originally tested, extensive tests of measure reliability were conducted. At present, no performance data for the electronic version of the measure are currently available. Below is the reliability testing summary for NQF #0438 Antithrombotic Therapy By End of Hospital Day Two, from which this measure is derived.

At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with monthly data transmission to The Joint Commission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of approximately 900 patient records identified at 30 pilot sites over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rates of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify potential measure enhancements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures.) The following is a list of the major tests conducted on the submitted data for this measure:

- Transmission of complete data
- Usage of data received (to understand if the healthcare organization provides the relevant service to treat the relevant population)
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity
- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors is as follows:

- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.

- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data element agreement rates reported to The Joint Commission for the chart-abstracted version of this measure for the time period of one year (4Q2010 – 3Q2011) have shown an overall agreement rate of 97.2%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-5.

Data Elements	total'n'numerator	total'n'denominator	rate
Antithrombotic Therapy Administered By End of Hospital Day 2	426	449	95.0%
Clinical Trial	712	714	99.7%
Comfort Measures Only	709	714	99.3%
Elective Carotid Intervention	711	714	99.6%
IV OR IA Thrombolytic Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival	464	476	97.5%
Reason for Not Administering Antithrombotic Therapy by End of Hospital Day 2	44	47	93.6%

These agreement rates are considered to be well within acceptable levels.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Agreement rates for individual data elements tested for the chart-abstracted version of this measure were within acceptable levels. Once data are available for analysis, it is expected that reliability tests of the eCQM version of this measure will yield similar results.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The Bonnie testing tool and environment were used to test the measure logic and value sets. Each data element and

logic statement was tested to confirm actual results met expectations. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure. An example of negative testing would be to include test cases with pediatric ages to ensure that pediatric patients are not included in the measure.

Denominator test cases positively test to ensure patients with principal diagnosis of ischemic stroke appropriately fall into the measure. We have created negative test cases, testing to ensure patients with a different principal diagnosis do not actually make it into the denominator.

Numerator test cases positively test to ensure patients receiving antithrombotic therapy administered the day of or day after hospital arrival pass the measure. Negative test cases ensure that a patient who does not receive antithrombotic therapy in the specified time frame do not pass the measure.

Denominator exclusion and exception test cases for this measure ensure that patients are properly removed from the denominator if they have length of stay less than 2 days, comfort measures, received intra-venous or intra-arterial thrombolytic therapy within specified timeframe, and medical reasons or patient refusal for not getting antithrombotic therapy. Negative test cases are also run, an example of this would be comfort measures that are ordered but not in the specified time frame. Testing confirmed patients meeting the exclusion and exception criteria are removed from the measure appropriately.

A review of the measure specifications was also conducted to confirm the logic was properly expressed within the current version of the QDM and confirmed the logic matches the clinical intent of the measure, as stated in the measure header. This is done through use of a logic checklist to facilitate review of the measure logic, according to several checks, a few examples of which are included below:

- Is the intent of the measure described in the measure description articulated/ captured in the measure logic?
- Do the logic elements map to definitions in the measure narrative, data dictionary or supporting reference documentation?
- Do the populations in the narrative align with the populations defined in the logic?
- Are the mathematic inequalities reflective of the measure intent and represent the intended populations (for example: when intended the inequality represents less than rather than less and or equal to)?

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Bonnie results provide coverage and passing rates. This measure reached 100% coverage, confirming there is a test case for each pathway of logic. This measure also has a 100% passing rate, confirming all test cases performed as expected.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

This measure performed as expected in Bonnie.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Exclusions in the eCQM align with the chart-based version of the measure, and are clinically necessary for the interpretation of the measure. As noted previously, the chart-abstracted version of this measure has been in national

use since the 4th quarter of 2009, and no data are available for the eCQM version of the measure. The below analysis addresses exclusions testing performed for the chart-abstracted version of the measure from which this measure is derived.

- The following exclusions were analyzed for frequency and variability across providers:
- Patients less than 18 years of age
- Patients who have a Duration of Stay less than 2 days
- Patient who have a Length of Stay greater than 120 days
- Patients with Comfort Measures documented on day of or day after arrival
- Patient enrolled in a Clinical Trial
- Patients admitted for Elective Carotid Intervention
- Patients discharged prior to the end of hospital day 2
- Patients with IV OR IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival
- Patients with a documented Reason for Not Administering Antithrombotic Therapy by End of Hospital Day 2

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are given below.

Exclusion: Comfort Measures Only documented on day of or day after arrival
Overall Occurrence n = 28,126
Overall Occurrence Percentage 4.46%
Minimum 0.31%
10th Percentile: 1.67%
Median: 4.40%
90th Percentile: 9.09%
Maximum: 30.8%

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%
10th Percentile: 0.24%
Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for Elective Carotid Intervention
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.5%
90th Percentile: 23.5%
Maximum: 95.2%

Exclusion: Patients with IV OR IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival

Overall Occurrence n = 25,354
Overall Occurrence Percentage 8.04%
Minimum: 0.25%
10th Percentile: 2.1%
Median: 6.72%
90th Percentile: 13.5%
Maximum: 47.4%

Exclusion: Patients with a documented Reason for Not Administering Antithrombotic Therapy by End of Hospital Day 2
Overall Occurrence n = 22,404
Overall Occurrence Percentage 7.11%
Minimum: 0.351%
10th Percentile: 1.48%
Median: 4.16%
90th Percentile: 14.9%
Maximum: 76.9%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Analysis of these data for the chart-asbtracted progenitor of this measure indicated that all exclusions were appropriate. It is believed that results for exclusions in the eCQM will be similar when sufficient data have been received to perform such an analysis.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

- ☒ **No risk adjustment or stratification**
- ☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
- ☐ **Stratification by** Click here to enter number of categories **risk categories**
- ☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of an HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

A similar methodology will be used for the eCQMs.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

NQF#0438: STK-5 Distribution of Measure Results
Quarter #hospitals Mean SD 90th% 75th% 50th% 25th% 10th Percentile

3Q2011	154	0.97045	0.0735	1	1	1	0.9697	0.92857
2Q2011	155	0.96839	0.08678	1	1	1	0.97143	0.93333
1Q2011	159	0.95328	0.12687	1	1	1	0.95833	0.91304
4Q2010	135	0.96697	0.104	1	1	1	0.96667	0.93023
3Q2010	129	0.95443	0.12216	1	1	1	0.96226	0.88889
2Q2010	122	0.95333	0.12219	1	1	1	0.95556	0.90625
1Q2010	100	0.9521	0.12692	1	1	1	0.96	0.87298
4Q2009	49	0.95947	0.10574	1	1	0.98519	0.96875	0.93478

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

It should be noted that since data collection on this measure is completely voluntary for both The Joint Commission and PCNASR, the hospitals reporting on this measure are self-selected, and therefore, presumably have a particular interest and concern for improving stroke care. For this reason, the measure results demonstrated by this group most likely significantly overstates the rate for the population of all health care organizations.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
***If only one set of specifications, this section can be skipped*.**

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Multiple data sources are not used for this measure.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Data not present in the structured field from which the measure draws will not be included in the measure calculation. In the Bonnie testing environment, missing data are tested as an expected "Fail" for that data element, and actual performance (whether or not the case fails) is compared to expectations to ensure missing data impacts data element and overall measure calculation as expected. This is the extent of missing data analysis that can be performed in Bonnie.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

See response for question 2b7.1 above.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

The measure performed as expected in the Bonnie testing environment.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment **Attachment:** STK5_eCQM_NQF_Measure_Feasibility_Assessment_Report.docx

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Not applicable.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
Value sets are housed in the Value Set Authority Center (VSAC), which is provided by the National Library of Medicine (NLM), in coordination with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

Viewing or downloading value sets requires a free Unified Medical Language System® (UMLS) Metathesaurus License, due to usage restrictions on some of the codes included in the value sets. Individuals interested in accessing value set content can request a UMLS license at (https://uts.nlm.nih.gov/license.html)

There are no other fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program http://www.jointcommission.org/ EHR Incentive Program https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor

- Purpose
- Geographic area and number and percentage of accountable entities and patients included

- Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
- Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 hospitals (2015)

- Name of program and sponsor: Hospital Accreditation Program; The Joint Commission
- Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

- Name of program and sponsor: EHR Incentive Program; Centers for Medicare & Medicaid Services
- Purpose: The American Recovery and Reinvestment Act of 2009 (ARRA) (Pub.L. 111–5) was enacted on February 17, 2009. Title IV of Division B of ARRA amends Titles XVIII and XIX of the Social Security Act (the Act) by establishing incentive payments to eligible professionals (EPs), eligible hospitals, and critical access hospitals (CAHs), and Medicare Advantage Organizations to promote the adoption and meaningful use of interoperable health information technology (HIT) and qualified electronic health records (EHRs). These incentive payments are part of a broader effort under the HITECH Act to accelerate the adoption of HIT and utilization of qualified EHRs.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4,847 active registrations (September, 2015)

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable.

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not applicable.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

While the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are currently available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs, one of which is this measure. Data will be accepted through February 28th, 2017.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

4b.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of

initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
Not applicable.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Not applicable.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0068 : Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet
0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy
0435 : STK 02: Discharged on Antithrombotic Therapy
0438 : STK 05: Antithrombotic Therapy By End of Hospital Day Two

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy – American Academy of Neurology

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Measures 0435 Discharged on Antithrombotic Therapy is the second (STK-2) measure in The Joint Commission stroke core measure set and also targets the ischemic stroke population; however, the timeframe for antithrombotic administration is different in this measure than STK-5. STK-2 focuses on hospital discharge and the prescription of antithrombotic medications at that time. All common data elements are completely aligned between the two measures. NQF#0438: STK 05: Antithrombotic Therapy By End of Hospital Day Two: The measures are completely harmonized to the extent possible, given the fact that the data source for #0438 is the paper medical record, and the data source for #2835 is the electronic health record. Measure 0068 is a physician performance measure and thus a different level of measurement. Measure 0068 encompasses a different target population, specifically patients with ischemic vascular disease who were discharged alive for acute myocardial infarction (AMI), coronary artery bypass graft (CABG) or percutaneous coronary interventions (PCI). Both of these measures evaluate physician practice as opposed to hospital processes.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not applicable.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Ann, Watt, awatt@jointcommission.org, 630-792-5944-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Lisa, Anderson, landerson2@jointcommission.org, 630-792-5008-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel (TAP) is to provide advisory oversight in literature review, measure content and maintenance of the specifications.
Members are:

Harold P. Adams, Jr., MD
University of Iowa Health Care
Iowa City, IA

Mark J. Alberts, MD
University of Texas Southwestern
Dallas, TX

Anne W. Alexandrov, RN
Health Outcomes Institute
Fountain Hills, AZ

Nadine Allyn, RD
American Heart Association
Dallas, Texas

Mary G. George, MD
Centers for Disease Control and Prevention
Atlanta, GA

Martin Gizzi, MD
Legacy Health System
Portland, OR

Marianna Gorbaty
Mathematica Policy Research, Inc.
Washington, DC

Judith Hinchey, MD
Tuft's Medical School
Steward St. Elizabeth's Medical Center
Boston, MA

Edward C. Jauch, MD
Medical University of South Carolina
Charleston, SC 2946

Irene Katzan, MD
Cleveland Clinic
Cleveland, OH

Kimberly E. Levasseur-Franklin, PharmD
Northwestern Memorial Hospital
Chicago, IL

Kathy Morrison, RN
Penn State Hershey Medical Center
Hershey, PA

Marilyn M. Rymer, MD
University of Kansas
Kansas City, MO

Jeffrey Saver, MD
UCLA Medical Center
Los Angeles, CA

Lee H. Schwamm, MD
Harvard Medical School
Massachusetts General Hospital
Boston, MA

Deborah Summers, RN
St. Luke's Brain and Stroke Institute
Kansas City, MO

Osama O. Zaidat, MD
St.Vincent Mercy Hospital
Toledo, OH

Richard D. Zorowitz, MD
Medstar National Rehabilitation Network
Washington, DC

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2012

Ad.3 **Month and Year of most recent revision:** 04, 2015

Ad.4 **What is your frequency for review/update of this measure?** Annual

Ad.5 **When is the next scheduled review/update for this measure?** 04, 2016

Ad.6 **Copyright statement:** Measure specifications are in the Public Domain

LOINC(R) is a registered trademark of the Regenstrief Institute.

Ad.7 **Disclaimers:** These performance measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. The measures and specifications are provided without warranty.

Ad.8 **Additional Information/Comments:** Not applicable.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2836
Measure Title: STK-06: Discharged on Statin Medication
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of ischemic stroke patients who are prescribed a statin medication at hospital discharge.

This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-4: Thrombolytic Therapy, STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs. STK-6, Discharged on Statin Medication, is one of six of the measures in this set that have been reengineered as eCQMs and are included in the EHR Incentive Program and Hospital Inpatient Quality Reporting Program.

Developer Rationale:
Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Recent guidelines recommend high-intensity statin therapy for patients with ischemic stroke unless contraindicated or patient characteristics require dosage modification.

Healthcare organizations that track this measure for internal quality improvement purposes have seen an increase in the measure rate over time. The chart-abstracted measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

Numerator Statement: Patients prescribed statin medication at hospital discharge.
Denominator Statement: Patients with a principal diagnosis of ischemic stroke.
Denominator Exclusions: Denominator Exclusions:
Patients admitted for elective carotid intervention. This exclusion is implicitly modeled by only including non-elective hospitalizations.
Patients with comfort measures documented.
Patients discharged to another hospital
Patients who left against medical advice
Patients who expired
Patients discharged to home for hospice care
Patients discharged to a health care facility for hospice care
Patients with an LDL-c of less than 70 mg/dL <30 days prior to arrival or any time during the hospital stay

Denominator Exceptions:
Patients with a reason for not prescribing statin medication at discharge.

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Pharmacy
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary or **Summary of prior review in 2012**

- The body of evidence provided supports that statin therapy should be prescribed for stroke prevention in patients with a prior history of ischemic stroke and transient ischemic attack . Based on these findings, there is a correlation between the use of statins in lowering LDL cholesterol and stoke occurrence. These findings are consistent across several clinical trials.
- 2010 AHA/ASA Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack, page 233.
  - Class I, Level of Evidence B – Statin therapy with intensive lipid-lowering effects is recommended to reduce the risk of stroke and cardiovascular events among patients with ischemic stroke or TIA who have evidence of arthrosclerosis, and LDL-c level > 100 mg/dL, and who are without known CHD.
- The 2012 Committee expressed no concerns regarding the evidence underlying this measure.

Exception to evidence

N/A

Guidance from the Evidence Algorithm

Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: Moderate; Consistency high (Box 5) → Box 5b Moderate

Questions for the Committee:

- *The developer attests the underlying evidence for the measure #0439 (chart-abstracted version) has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the chart-abstracted measure has not changed and can be applied to the eMeasure, therefore there is no need for repeat discussion and voting on Evidence?*

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer states no performance data for the electronic version of this measure are yet available.
- This eCQM is currently included in the Hospital Inpatient Quality Reporting Program (IQR) and EHR Incentive program.
  - In CY2016, CMS is requiring organizations participating in HIQR to electronically submit one quarter of data on 4 of 28 available eCQMs. This measure is one of the 28 eCQMs.

- The developer provided a summary of data from the literature that demonstrated underutilization of antithrombotic therapy and a performance gap in Medicare patients.
- The developer also provided performance data from the chart abstracted measure which seems to be very high with little to no room for improvement.

Disparities

- The developer does not provide disparities data from use of this measure.
- Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age."

Questions for the Committee:

- *Is there a gap in care that warrants a national performance eMeasure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **same issue with evidence as standard measure; changed criteria without new evidence

**This measure is a process measure the evidence of which seems to relate directly to the outcome being measured. This is not a maintenance measure nor is it measuring a health outcome or PRO.

**chart-abstracted measure evidence can be applied to emeasure

1b. Performance Gap

Comments: **performance is high, so not much gap; theoretical gap for women and minorities, but not shown in data from this measure (not available yet), nor from the traditional version of measure

**Performance data on the measure was provided which seem to demonstrate a gap in care in which demonstrated improvement over time which should warrant a national performance measure. It did not measured by population subgroups so it did not demonstrate disparities in care other than clinically but did attach a summary of data that favorably addresses the issue of disparities.

**performance data provided; no disparities info

1c. High Priority (previously referred to as High Impact)

Comments: **Not applicable

**nA

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic Clinical Data, Electronic Health Record (EHR), Imaging/Diagnostic Study, Laboratory Data, and Pharmacy Data. This is an eMeasure.

Specifications:

- HQMF specifications for eMeasure are included in the document set on SharePoint. See eMeasure Technical Review below.
- The developer states that changes have been made to the eCQM specifications to reflect revisions to the chart abstracted measure from which this measure is derived (NQF measure #0439).

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor review:

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously Demonstrated through Bonnie testing
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible. This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☒ **No**

Method(s) of reliability testing:

- Dataset used for testing included 28 synthetic records created in the BONNIE testing system. Measure not tested in EHR.
- Data element validity testing was performed and will count for data element reliability as well – see validity testing section.

Guidance from the Reliability Algorithm Precise specifications (Box 1) →empiric reliability testing as specified (Box 2) → empiric validity testing of patient-level data (Box 3) → YES: Use rating from validity testing of patient-level data elements (Box 10)→Assessed the measure logic only (Box 11) →Low certainty or confidence data used in measure are

valid (Box 12b)→Low (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Questions for the Committee:

- o *Is the test sample adequate to generalize for widespread implementation?*
- o *Do the results from the Bonnie tool demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- The Bonnie testing tool and environment with 28 synthetic patient records were used to test the measure logic and value sets. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure.
- The numerator, denominator, exclusions/exceptions, and logic statement were tested to confirm actual results met expectations.
- The developer also conducted a review of the measure specifications to confirm the measure logic was properly expressed with the current version of the QDM and that the logic matches the clinical intent of the measure.

Validity testing results:

- The testing results from the Bonnie tool reached 100% coverage and confirmed there was a test case for each pathway of logic (negative and positive test cases).
- The measure also had a 100% passing rate which confirmed that all the test cases performed as expected.
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer submitted additional information on exclusions and meaningful differences.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Is it reasonable to assume that the impact of exclusions will be similar for the eMeasure and the chart-abstracted measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer submitted additional information on exclusions and meaningful differences.

Question for the Committee:

- *For meaningful differences in the chart-abstracted measure the developer refers to the performance scores over time. Does the SC think the eMeasure will demonstrate similar results to the chart-abstracted measure which show little to no room for improvement?*

2b6. Comparability of data sources/methods: Not Applicable

2b7. Missing Data

- The developer describes the handling of missing data but does not provide information on the frequency of missing data or potential impact on measure results due to limited testing abilities using the Bonnie tool.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → threats to validity empirically assessed to some extent (Box2) → empirical testing of data elements using BONNIE tool (Box 3,10) →Moderate

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **The issue with validity rests with the fact that this measure is intended for electronic submission but, to date, there is no performance data on such submissions. The validity on the measure itself seems clear.

**No concerns

2a2. Reliability Testing

Comments: **only done with minimal bonnie data

**There is a lack of validity data with respect to electronic submission

**Data element testing assessed

2b2. Validity Testing

Comments: **no testing in real EHRs so not so clear?

**Only a lack of specific performance testing of electronic submission

**agree insufficient rating for validity

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **The minimum requirement is testing in EHR systems from more than 1 EHR vendor. Developers should test on the number of EHRs they feel appropriate. It is highly desirable that measures are tested in systems from multiple vendors.

**Again, there is a lack of reliability data to respect to electronic submission.

**testing only done at data element level; low reliability

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is EHR; all data elements are in defined fields in EHR.
- The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible.
- This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment.
- Per developer, they were unable to directly assess data accuracy of the data elements in an EHR system. Workflow, or the degree to which the data elements are captured during the course of care was also not assessed.

Questions for the Committee:

- *Is it likely the required data elements are routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **feasibility is partially justified by this measure being used to submit data in 2015; so why could not these 2015 data be used for other testing?

**No basis for evaluation

**Data available in EHR however how they are collected during care not assessed

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- CMS Hospital Inpatient Quality Reporting Program (IQR)/EHR Incentive Program: The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. The EHR Incentive Program provides incentive payments to promote the adoption and meaningful use of interoperable HIT and qualified EHRs.
- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results:

- The developer states that while the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are currently available.
- In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs, one of which is this measure. Data will be accepted through February 28th, 2017.

Unexpected findings (positive or negative) during implementation:

- Developer did not identify unexpected findings during implementation.

Potential harms:

- Developer did not identify any unintended consequences related to this measure.

***Questions for the Committee**:*

- o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **CMS using currently

**The only unintended consequences that persons not in need of a statin may be prescribed same however, the advantage of statin administration posted ischemic stroke is well-documented and the measure excludes persons there is a documented reason for not prescribing statins.

**Not publicly reported but in CMS accountability program

Criterion 5: Related and Competing Measures

Related or competing measures:

- 0074 : Chronic Stable Coronary Artery Disease: Lipid Control
- 0118 : Anti-Lipid Treatment Discharge
- 0439 : STK-06: Discharged on Statin Medication
- 1519 : Statin Therapy at Discharge after Lower Extremity Bypass (LEB)
- 0545 : Adherence to Statins for Individuals with Diabetes Mellitus

- 0543 : Adherence to Statin Therapy for Individuals with Cardiovascular Disease – CMS – no longer NQF endorsed
- 0547 : Diabetes and Medication Possession Ratio for Statin Therapy – CMS – no longer NQF endorsed
- 0639 : Statin Prescribed at Discharge – CMS – no longer NQF endorsed

Harmonization:

- #0074, #0118, and #1519 target diagnoses other than ischemic stroke or specific surgical procedures. In addition, #0074 and #0118 are provider level measures.
- #0439 and #2836 are completely harmonized to the extent possible, given the fact that the data source for #0439 is the paper medical record and the data source for #2836 is the electronic health record.
- The developer states that #0545 is not in the NQF database and does not address harmonization; however, endorsement was last renewed September 2, 2014.

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0439 NQF Project: Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The focus of the measure is to reduce patient morbidity and mortality through modification of risk factors for stroke, specifically through the use of statin medications to reduce low-density lipoprotein cholesterol (LDL-c). Statin prescribed at discharge >> decreased LDL-c >> secondary stroke prevention >>decreased morbidity and mortality.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
This measure is consistent with the guidelines recommended by the American Heart Association/American Stroke Association for the prevention of stroke in patients with stroke and transient ischemic attack. An elevated serum lipid level has been a well-documented risk factor for coronary artery disease (CAD). Recently, there has been an increased focus on examining the relationship between elevated lipid levels and the incidence of stroke. In particular, some recent clinical trials have analyzed the association between lipids and non-hemorrhagic stroke. The reduction of LDL cholesterol, through lifestyle modification and drug therapy, for the prevention of strokes and other vascular events is recommended for patients with CAD in the National Cholesterol Education Program Adult Treatment Panel III (NCEP ATP III) Guidelines. In addition, recent evidence from the Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) trial supports the use of statins to lower LDL cholesterol in stroke patients without prior CAD and a fasting LDL >/= 100 mg/dL.

Based on these guidelines, all patients with ischemic stroke should have lipid profile measurement performed within 48 hours of admission unless outpatient results are available from within the past 30 days. Treatment for secondary prevention should be initiated in patients who meet NCEP ATP III criteria in the presence of LDL >/= 100 mg/dL, or continued for patients who were previously on lipid-lowering therapy and have an LDL< 100 mg/dL.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: The number of studies supporting the body

of evidence for statin therapy is in the thousands. A recent editorial by Lori Mosca, MD, MPH, PhD published in Journal of the American College of Cardiology (2012) noted more than 2,300 potential studies in the literature related to sex-specific outcomes and the effects of statins on the prevention of cardiovascular disease.

The Third Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III, or ATP III) extensively analyzed the results of 63 clinical trials. Findings from these studies strongly influenced the development of ATP III recommendations (2002).

To date, the single most important study of statin therapy and stroke incidence is The Stroke Prevention by Aggressive Reduction of Cholesterol Levels (SPARCL) trial. SPARCL randomly assigned 4731 patients who had had a stroke or TIA within one to six months before study entry, had low-density lipoprotein (LDL) cholesterol levels of 100 mg to 190 mg per deciliter (2.6 to 4.9 mmol per liter) and had no known history of coronary heart disease to double-blind treatment with 80 mg of atorvastatin per day or placebo. The incidence of fatal or non-fatal stroke in patients receiving atorvastatin was reduced by 16%. SPARCL is unique in that it is the only trial to study a cohort of patients with no known history of coronary heart disease.

A Cochrane Database of Systematic Review (10 August 2011) identified eight randomized control trials (RCTs) involving 625 participants. The review included all RCTs comparing statins (any type and dosage) versus placebo or no treatment, administered within two weeks of the onset of acute ischemic stroke or TIA. The search strategy included the Cochrane Stroke Group's Trials Register (November 2010); the Cochrane Central Register of Controlled Trials (CENTRAL)(The Cochrane Library 2010, Issue 4): MEDLINE (1950 to November 2010); and EMBASE (1980 to November 2010). In addition, ongoing trials and research registers (November 2010) were also searched and reference lists from relevant articles and contacted authors checked to further identify published, unpublished, and ongoing trials. Two review authors independently selected studies for inclusion and extracted data.
A MEDLINE search, using PubMed, of all literature related to "Hydroxymethylglutaryl-CoA Reductase Inhibitors/therapeutic use" was done for the years 2007 to 2012 and yielded 240 results. A meta-analysis of 12 studies (Biffi, et al., 2011), comprising 2013 statin users and 9682 non-users, investigated the association between prestroke statin use and clinical outcome. A meta-analysis of 18 RCTs (Kostis WJ, et al., 2012) of statins with sex-specific outcomes (N=141,235 participants, 40,275 2omen, 21,468 cardiovascular events, i.e., coronary heart disease and stroke) demonstrated the benefit of statins in decreasing morbid and mortal cardiovascular events in apparently healthy individual and in those with clinically evident cardiovascular disease.

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting the effectiveness of statin therapy on stroke outcome is moderate. There is evidence that stain therapy in both primary and secondary prevention significantly reduces subsequent major coronary events but only marginally reduces the risk of stroke recurrence. There is no clear evidence of beneficial effect from statins in those with previous hemorrhagic stroke, and it is unclear if statins should be started immediately post stroke or later (Feher, et al., 2010).

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of evidence consistently supports that stroke and TIA patients should be prescribed statin therapy for stroke prevention. No position against the importance of statin therapy for secondary stroke prevention was identified in the literature.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
Statins play an important role in brain ischemia. These drugs reduce cholesterol levels, which have been related to a reduction in vascular event risk, and also have other functions besides cholesterol metabolism, called pleiotropic effects. Statins play an important role during the acute phase of ischemia, and might have neuroprotective effects, as they act in several mechanisms during the acute phase of stroke, such as nitric oxide (NO) and glutamate metabolism, inflammation, platelet aggregation, immune responses, and apoptosis. They also have other functions which can be related, with better long-term outcome, to neurorepair mechanisms. Statins promote angiogenesis, endogenous cell proliferation, neurogenesis and new synapse formation (Rodriguez-Yáñez M, et al., 2008).

The potential positive action of stains during acute cerebrovascular ischemic event are two-fold: a neuroprotective effect, limiting damage and improving recovery; and a preventative effect on early recurrence. The SPARCL trial did report a slight increase in hemorrhagic stroke (n=55) for the atorvastatin group compared to placebo (n=33). Based on a Cochrane Review of eight RCTs of statins, no patients died from ischemic stroke or from adverse drug effects, bleeding or infections among 444 participants in six studies

where these outcomes were reported. Statin treatment did not reduce all-cause mortality compared with placebo or no treatment (OR 1.51, 95% CI 0.60 60 3.81) in the 431 patients enrolled in seven studies. No cases of rhabdomyolysis occurred in 274 patients enrolled in three studies.

Diet, exercise, and lifestyle modification are assumed to be cost-effective stroke prevention strategies; however, no specific studies on the cost-effectiveness of statin therapy for the prevention of stroke recurrence were noted in the literature. Most of the literature focuses on the primary prevention of cardiovascular disease. According to Chan and colleagues (2007) recent clinical trials have found that high-dose statin therapy, compared with conventional-dose statin therapy, reduces the risk of cardiovascular events in patients with acute coronary syndromes (ACS) and stable coronary artery disease (CAD). However, the actual benefit and cost-effectiveness of high-dose statin therapy are unknown. The daily cost difference between a high- and conventional-dose statin would need to be <$1.70, $2.65, and $3.55 to yield incremental cost-effective ratios below $50,000, $100,000, and $150,000 per quality-adjusted life year (QALY).

Another study (Lazar LD, et al., 2011), evaluated the cost-effectiveness of statin therapy for primary prevention. This study utilized the Coronary Heart Disease (CHD) Policy Model, an established computer simulation, Markov state-transition model of CHD incidence, prevalence, mortality and costs in the US population >35 years of age. Cost-savings were projected for persons with LDL > 100 mg/dL and moderately high risk, LDL > 130 mg/dL and moderate risk, and LDL > 160 mg/dL and lower or lowest risk with the cost of statins at $4/month. Using this strategy, it would be possible to prevent 14,000 CHD deaths per year and save over $1.4 billion a year compared with current levels of treatment.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: National Cholesterol Education Program, Adult Treatment Panel (ATP) III (2002). During our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, and provided information that was accessible and met the requirements set out in this measure maintenance form.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: The Third Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III, or ATP III) presents the National Cholesterol Education Program's (NCEP's) updated recommendations for cholesterol testing and management.
ATP III contains both evidence statements and specific recommendations based on these statements. Each evidence statement is qualified according to category of evidence (A-D) and strength of evidence (1-3), as follows:

Type of Evidence
Category and Description of Type of Evidence
A – Major randomized controlled clinical trials (RCTs)
B – Smaller RCTs and meta-analyses of other clinical trials
C – Observational and metabolic studies
D – Clinical experience

Strength of Evidence
Category and Description of Strength of Evidence
1 – Very strong evidence
2 – Moderately strong evidence
3 – Strong trend

1c.13 Grade Assigned to the Body of Evidence: A1 (Major randomized controlled clinical trials (RCTs)/Very strong evidence) // B1 (Smaller RCTs and meta-analyses of other clinical trials/Very Strong evidence)

1c.14 Summary of Controversy/Contradictory Evidence: There is some recent debate about the use of statins in women who do not have heart disease but do have high levels of cholesterol. An observational study (Culver AL, et al., 2012), involving 153,840 postmenopausal women enrolled in the Women's Health Initiative, reported a 48% greater risk of developing diabetes mellitus for women taking statin medications when compared to women not taking the medication at baseline. However, a new meta-analysis

(Kostis WJ, et al., 2012) of 18 trials with 141,235 participants concluded that statin therapy was beneficial in both the primary and secondary prevention settings for both men and women.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- Adams RJ, Albers G, Alberts MJ, Benavente O, Furie K, Goldstein LB, Gorelick P, Halperin J, Harbaugh R, Johnston SC, Katzan I, Kelly-Hayes M, Kenton EJ, Marks M, Sacco RL, Schwamm LH. Update to the AHA/ASA recommendations for the prevention of stroke in patients with stroke and transient ischemic attack. Stroke. 2008;39(5):1650.
- Biffi A, Devan WJ, Anderson CD, Cortellini L, Furie KL, Rosand J, Rost NS. Statin treatment and functional outcome after ischemic stroke: case-control and meta-analysis. Stroke. 2011;42(5): 1314-9.
- Chan PS, Nallamothu BK, Gurm HS, Hayward RA, Vijan S. Incremental benefit and cost-effectiveness of high-dose statin therapy in high-risk patients with coronary artery disease. Circulation. 2007;115:2398-2409.
- Craig SR, Amin RV, Russell DW, Paradise NF. Blood cholesterol screening influence of fasting state on cholesterol results and management decisions. J Gen Intern Med. 2000 Jun;15(6):395-9.
- Culver AL, Ockene IS, Balasubramanian R, Olendzki BC, Sepavich DM, Wactawski-Wende J, Manson JE, Qiao Y, Liu S, Merriam PA, Rahilly-Tierny C, Thomas F, Berger JS, Ockene JK, Curb JD, Ma Y. Statin Use and Risk of Diabetes Mellitus in Postmenopausal Women in the Women´s Health Initiative.
Arch Intern Med. 2012; 172: 144 – 152.
- Feher A, Pusch G, Koltai K, Tibold A, Gasztonyi B, Szapary L, Feher G. Statintherapy in the primary and secondary prevention of ischaemic cerebrovascular diseases. Int J Cardiol. 2011; 148(2): 131-8.
- Feinberg WM, Albers GW, Barnett HJM, et al. Guidelines for the Management of Transient Ischemic Attacks. From the Ad Hoc Committee on Guidelines for the Management of Transient Ischemic Attacks of the Stroke Council of the American Heart Association. 1994.
- Gore JM, Goldberg RJ, Matsumoto AS, et al. Validity of serum total cholesterol level obtained within 24 hours of acute myocardial infarction. Am J Cardiol. 1984;54:722-725.
- Grundy SM, Cleeman JI, Merz CNB, Brewer, HB, et. al. Implications of Recent Clinical Trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. Circulation. 2004;110:227-239
- High-Dose Atorvastatin after Stroke or Transient Ischemic Attack. (New England Journal of Medicine. NEJM Vol. 355 2006:549-559.
- Kostis WJ, Cheng JQ, Dobrzynski BA, Cabrera J, Kostis JB. Meta-analysis of statin effects in women versus men. Journal of American College of Cardiology. 2012;59(6):572-82.
- Lazar LD, Pletcher MJ, Coxson PG, Bibbins-Domingo K, Goldman L. Cost-effectiveness of statin therapy for primary prevention in a low-cost statin era. Circulation. 200;124:146-153.
- Mitka M. Some queston use of statins to reduce cardiovascular risks in healthy women. JAMA. 2012;37(9):893-94.
- Pitt B, Loscalzo J, Ycas J, Raichlen JS. Lipid Levels After Acute Coronary Syndromes. J Am Coll Cardiol 2008;51;1440-1445.
- Rodrigues-Yáñez M, Agulla J, Rodrigues-González R, Sobrino, T, Castillo J. Statins and stroke. Ther Adv Cardiovasc Dis. 2008;2(3): 157-66.
- Sacco RL, Adams R, Albers G, Alberts MJ, Benavente O, Furie K, Goldstein LB, Gorelick P, Halperin J, Harbaugh R, Johnston SC, Katzan I, Kelly-Hayes M, Kenton EJ, Marks M, Schwamm LH, Tomsick T. Guidelines for Prevention of Stroke in Patients With Ischemic Stroke or Transient Ischemic Attack: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association Council on Stroke: Co-Sponsored by the Council on Cardiovascular Radiology and Intervention. Stroke. Vol. 37, 2006:577.
- Schellinger PD, Bryan RN, Caplan LR, Detre JA, et al. Evidence-based guideline: The role of diffusion and perfusion MRI for the diagnosis of acute ischemic stroke: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Neurology 2010;75;177-185.
- Squizzato A, Romualdi E, Dentali F, Ageno W. Statins for acute ischemic stroke. Cochrane Database Syst Rev. 2011 Aug 10;(8): CD007551.
- Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Final Report Circulation Vol. 106 2002: 3143-3421.
- Van Dis FJ, Keilson LM, Rundell CA, et al. Direct measurement of serum low-density lipoprotein cholesterol in patients with acute myocardial infarction on admission to the emergency room. Am J Cardiol. 1996;77:1232-1234.
- Weiss R, Harder M, Rowe J. The relationship between nonfasting and fasting lipid measurements in patients with or without type 2 diabetes mellitus receiving treatment with 3-hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors. Clin Ther. 2003 May;25(5):1490-7.

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)*:
2010 Update to the AHA/ASA Recommendations for the Prevention of Stroke in Patients with Stroke and Transient Ischemic Attack,

page 233.
Class I, Level B Recommendation
Statin therapy with intensive lipid-lowering effects is recommended to reduce the risk of stroke and cardiovascular events among patients with ischemic stroke or TIA who have evidence of atherosclerosis, and LDL-c level > 100 mg/dL, and who are without known CHD.

1c.17 Clinical Practice Guideline Citation: Furie KL, Kasner SE, Adams RJ, Albers GW, et al. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack. A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2011;42:232-33.

1c.18 National Guideline Clearinghouse or other URL: http://stroke.ahajournals.org/content/42/1/227.full.pdf

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: This guideline was approved by the American Heart Association Science Advisory and Coordinating Committee on July 13, 2010. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.23 Grade Assigned to the Recommendation: Class I, Level of Evidence A/B

1c.24 Rationale for Using this Guideline Over Others: The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: High 1c.26 Quality: Moderate1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0439_Evidence_MSF5.0_Data-635827722667216159.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with

significant costs. Recent guidelines recommend high-intensity statin therapy for patients with ischemic stroke unless contraindicated or patient characteristics require dosage modification.

Healthcare organizations that track this measure for internal quality improvement purposes have seen an increase in the measure rate over time. The chart-abstracted measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This measure is a legacy eCQM that is currently included in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data for the electronic version of the measure are yet available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs. This measure is one of the 28.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
In October, 2009, The Joint Commission added the chart-abstracted stroke (STK) measure set as a new core measure option to meet performance measure requirements for Joint Commission hospital accreditation purposes. Below is the specified level of analysis for STK-6 beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 979 denominator cases; 876 numerator cases; 48 hospitals; 0.89479 national aggregate rate; 0.88567 mean of hospital rates; 0.18656 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.94949 50th percentile rate/median rate; 0.85165 25th percentile rate/lower quartile; and, 0.66667 10th percentile rate.

CY 2010: 12,042 denominator cases; 11,188 numerator cases; 137 hospitals; 0.92908 national aggregate rate; 0.89366 mean of hospital rates; 0.12881 standard deviation; 1.0 90th percentile rate; 0.97059 75th percentile rate/upper quartile; 0.94231 50th percentile rate/median rate; 0.86667 25th percentile rate/lower quartile; and, 0.70647 10th percentile rate.

CY 2011: 18,608 denominator cases; 17,512 numerator cases; 157 hospitals; 0.9411 national aggregate rate; 0.90293 mean of hospital rates; 0.12829 standard deviation; 1.0 90th percentile rate; 0.97297 75th percentile rate/upper quartile; 0.94361 50th percentile rate/median rate; 0.88889 25th percentile rate/lower quartile; and, 0.78571 10th percentile rate.

CY 2012: 18,998 denominator cases; 18,165 numerator cases; 157 hospitals; 0.95615 national aggregate rate; 0.9412 mean of hospital rates; 0.08087 standard deviation; 1.0 90th percentile rate; 0.99371 75th percentile rate/upper quartile; 0.96825 50th percentile rate/median rate; 0.925 25th percentile rate/lower quartile; and, 0.84375 10th percentile rate.

CY 2013: 29,350 denominator cases; 28,374 numerator cases; 262 hospitals; 0.96675 national aggregate rate; 0.9469 mean of hospital rates; 0.09876 standard deviation; 1.0 90th percentile rate; 0.99649 75th percentile rate/upper quartile; 0.97938 50th percentile rate/median rate; 0.94444 25th percentile rate/lower quartile; and, 0.86585 10th percentile rate.

CY 2014: 140,296 denominator cases; 136,545 numerator cases; 1296 hospitals; 0.97326 national aggregate rate; 0.95761 mean of hospital rates; 0.08458 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.98507 50th percentile rate/median rate; 0.95303 25th percentile rate/lower quartile; and, 0.89474 10th percentile rate.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al., 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

In the national REasons for Geographic and Racial Differences in Stroke (REGARDS) cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years of age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age : RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer,et al., 2010).

In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).

Differences in access to and use of stroke prevention therapies in all racial and ethnic groups has been poorly documented and understudied; however, one study of 5840 stroke survivors as part of the National Health Interview Survey, found that women, blacks

or African Americans, and the poor were significantly less likely to fill prescriptions because of costs. Disparities are reduced when patients have health insurance and ready access to care.

According to the literature, minorities are less likely to receive medications for secondary prevention, including statin therapy for hyperlipidemia (Cruz-Flores, et al., 2011). Yood and colleagues (2006) found that blacks or African Americans newly diagnosed with dyslipidemia and prescribed statins were 36% less likely to achieve low-density lipoprotein goals over time (hazard ration 0.64, 95% CI O.61 to 0.68). This disparity persisted after low-density lipoprotein testing and adjustment for statin adherence (hazard ration 0.60, 95% CI 0.57 to0.63). Another study by Mark and associates (2007) noted that blacks or African Americans were less likely than whites to be switched between lipid-lowering agents (OR 0.68 95% CI 0.60 to 0.78), to have treatment adjusted (OR 0.53, 95% CI 0.43 to 0.66), or to be prescribed higher medication dosages (OR 0.75, 95% CI 0.67 to 0.84).

Since the last endorsement date, Schwamm and colleagues (2010) found that black patients with stroke received fewer evidence-based care processes than Hispanic or white patients. Lipid-lowering therapy at discharge for patients with low-density lipoprotein (LDL) > 100, or those on lipid-lowering agents before hospital admission, or in whom LDL was not measured in the past 30 days, was one care process evaluated in this study. Blacks had lower odds relative to white patients of receiving lipid therapy at discharge after adjustment for both patient and hospital level variables: OR 0.91 [95% CI 0.88-0.96]

A more recently published study (Qian F, et al, 2013) from GWTG also noted statistically significant (P<0.01) racial and ethnic disparities for lipid-lowering therapy at discharge. Using patient data (n=200,900) from the American Heart Association/American Stroke Association Get With The Guidelines (GWTG)-Stroke program from April 2003 through December 2008, the following performance measure rates were reported for lipid-lowering medication prescribed at discharged : non-Hispanic White (n=170,694) 84.3%; non-Hispanic Black (n=20,514) 83.8%; Hispanic (n=6632) 85.8%; and non-Hispanic Asian American (n=3060) 86.7%.

2014 data from the Paul Coverdell National Acute Stroke Registry (PCNASR) reported greater disparity for women (95.9%) compared to men (97.0%) prescribed statin therapy at discharge. Similar rates of statin therapy prescribed at discharge were reported for white patients (96.3%) and other races (96.7%); aggregate rate 96.4%; n=45,249 (Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, 2014).

• Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, Yan T, Escarce JJ. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-8.
• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, Griffith P, Gorelick PB, Howard G, Leira EC, Morgenstern LB, Ovbiagele B, Peterson E, Rosamond W, Trimble B, Valderrama AL, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• Gillum RF,Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 2011;42(11):3294-6.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-75.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Karve S, Balkrishnan R, Seiber E Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Mark TL, Axelsen KJ, Mucha L, Sadkova Y. Racial differences in switching, augmentation, and titration of lipid-lowering agents by Medicare/Medicaid dual-eligible patients. Am J Manag Care. 2007;13(3):S72-S79.
• Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, Kerber KA, Skolarus LE, Meurer WJ, Burke JF, Adelman EE, Baek J, Lisabeth LD. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol. 2013;74:778–785.

• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Yood MU, McCarthy BD, Kempf J, Kucera GP, Wells K, Oliveria S, Stang P. Racial differences in reaching target low-density lipoprotein goal among individuals treated with statin therapy. Am Heart J. 2006;152:777-784.
• Schwamm LH, Syed FA, Reeves MJ, Smith EE, Saver JL, Messe S, Bhatt DL, Grau-Sepulveda MV, Peterson ED, Fonarow GC. Temporal trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With the Guidelines-Stroke hospitals. Circ Cardiovasc Quality Outcomes. 2013;6:543-549.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality, Severity of illness

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States remains attributable to stroke. More women than men die of stroke each year. Women accounted for almost 60% of US stroke deaths in 2008 (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, et al., 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 (P<0.05).168 (US Burden of Disease Collaborators, 2013) . Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

According to the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III/ATP III) (National Institutes of Health, 2002), statins should be considered as first-line drugs when LDL-lowering drugs are indicated to achieve LDL treatment goals. HMG CoA reductase inhibitors (statins) are powerful LDL-lowering drugs. Statin therapy reduces the risk of acute coronary syndromes, coronary procedures, and other coronary outcomes in both primary and secondary prevention. It also reduces the risk of stroke in secondary prevention.

The Stroke Prevention by Aggressive Reduction of Cholesterol Levels (SPARCL) trial (Amarenco, P, et al., 2006) concluded that in patients with recent stroke or TIA and without known coronary heart disease, 80 mg of atorvastatin per day reduced the overall incidence of strokes and cardiovascular events, despite a small incidence of hemorrhagic stroke. The trial randomly assigned 4731 patients who had had a stroke or TIA within one to six months before study entry, had low-density lipoprotein (LDL) cholesterol levels

of 100 to 190 mg per deciliter (2.6 to 4.9 mmol per liter), and had no known coronary heart disease to double-blind treatment with 80 mg of atorvastatin per day or placebo. The primary end point was a first nonfatal or fatal stroke. The mean LDL cholestrerol level during the trial was 73 mg per deciliter (1.9 mmol per liter) among patients receiving atorvastatin and 129 mg per deciliter (3.3 mmol per liter) among patients receiving placebo. During a median follow-up of 4.9 years, 265 patients (11.2 percent) receiving atorvastatin and 311 patients (13.1 percent) receiving placebo had a fatal or nonfatal stroke (5-year absolute reduction in risk, 2.2 percent; adjusted hazard ratio, 0.84; 95 percent confidence interval, 0.71 to 0.99; P=0.03; unadjusted P=0.05) The atorvastatin group had 218 ischemic strokes and 55 hemorrhagic strokes. The five-year absolute reduction in the risk of major cardiovascular events was 3.5 percent (hazard ratio, 0.80; 95 percent confidence interval, 0.69 to 0.92; P=0.002). The overall mortality rate was similar, with 216 deaths in the atorvastatin group and 211 deaths in the placebo group (P=0.98), as were the rates of serious adverse events. Elevated liver enzyme values were more common in patients taking atorvastatin.

In 2013, the American College of Cardiology (ACC) / American Heart Association (AHA) updated the Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults, and strongly recommended high-intensity statin therapy for secondary prevention in patients with clinical atherosclerotic cardiovascular disease (ASCVD), unless contraindicated. Patients with ischemic stroke due to atherosclerosis are included in this first of four statin benefit groups. High-intensity statin therapy should be initiated or continued as first-line therapy in both women and men =75 years. When high-intensity statin therapy is contraindicated or for those patients unable to tolerate high-intensity statin therapy, moderate-intensity statin therapy should be used as a second option. High-intensity therapy may also be reasonable for ischemic stroke patients > 75 years, if risk-reduction benefits outweigh the risk of adverse events and the patient can tolerate it (Stone NJ, et al, 2013). AHA/ASA Guidelines for the Prevention of Stroke in Patients with Stroke or Transient Ischemic Attack (Kernan WN, et al, 2014) were revised the following year to align with 2013 ACC/AHA recommendations. Statin therapy was recommended to reduce the risk of stroke and cardiovascular events for patients with ischemic stroke or TIA presumed to be of atherosclerotic origin and LDL > 100 with or without evidence for other ASCVD, as well as, patients with LDL < 100 and no evidence for other clinical ASCVD.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

• Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• George, M., Xin, Tong, McGruder, H., Yoon, P., Rosamond, W., Winquist, A., Hinchey, J., Wall, H., Pandey, D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Grundy SM, Cleeman JI, Merz CNB, Brewer, HB, et. al. Implications of Recent Clinical Trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. Circulation. 2004;110:227-239
• High-Dose Atorvastatin after Stroke or Transient Ischemic Attack. (New England Journal of Medicine. NEJM Vol. 355 2006:549-559.
• Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, Fang MC, Fisher M, Furie KL, Heck DV, Johnston SC, Kasner SE, Kittner SJ, mitchell PH, Rich MW, Richardson D, Schwamm LH, Wilson JA. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2014;45:241-2160-2236.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Stone NJ, Robinson J, Lichtenstein AH, Noel Bairey Merz C, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, McBride P, Schwartz JS, Shero Jr, ST, Smith SC, Watson K, Wilson PWF. "Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Disease in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. [In eng]. Circulation 11, (Nov 2013): 1-84.
• Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Final Report Circulation Vol. 106 2002: 3143-3421.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/ecqm_library.html

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is an eMeasure **Attachment:** STK-6_MAT.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** DischargedonStatinMedication_v4_Wed_Apr_01_12.18.50_CDT_2015.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Changes have been made to the eCQM specifications in order to reflect revisions to the chart abstracted measure from which this measure is derived.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients prescribed statin medication at hospital discharge.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Statin Medication
- Statin Medication is represented with the QDM datatype and value set of Medication, Discharge: Statin (OID: 2.16.840.1.113883.3.117.1.7.1.225)

Non-Elective Inpatient Encounter
- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients with a principal diagnosis of ischemic stroke.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Principal Diagnosis of Ischemic Stroke

- Ischemic Stroke is represented with the QDM datatype and value set of Diagnosis, Active: Ischemic Stroke (OID: 2.16.840.1.113883.3.117.1.7.1.247)
- Ordinality: Principal (OID: 2.16.840.1.113883.3.117.1.7.1.14)

Non-Elective Inpatient Encounter

- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Denominator Exclusions:
Patients admitted for elective carotid intervention. This exclusion is implicitly modeled by only including non-elective hospitalizations.
Patients with comfort measures documented.
Patients discharged to another hospital
Patients who left against medical advice
Patients who expired
Patients discharged to home for hospice care
Patients discharged to a health care facility for hospice care
Patients with an LDL-c of less than 70 mg/dL <30 days prior to arrival or any time during the hospital stay

Denominator Exceptions:
Patients with a reason for not prescribing statin medication at discharge.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Denominator Exclusion Data Elements:

Non-Elective Inpatient Encounter

- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

Discharge Status (modeled as Attributes of the above Non-Elective Inpatient Encounter)

- Discharge status: Left Against Medical Advice (OID: 2.16.840.1.113883.3.117.1.7.1.308)
- Discharge status: Patient Expired (OID: 2.16.840.1.113883.3.117.1.7.1.309)
- Discharge status: Discharge To Acute Care Facility (OID: 2.16.840.1.113883.3.117.1.7.1.87)
- Discharge status: Discharged to Home for Hospice Care (OID: 2.16.840.1.113883.3.117.1.7.1.209)
- Discharge status: Discharged to Health Care Facility for Hospice Care (OID: 2.16.840.1.113883.3.117.1.7.1.207)

Comfort Measures

- Comfort Measures are represented with the QDM datatypes and value set of:
- Intervention, Order: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)
- Intervention, Performed: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)

Emergency Department Visit

- Emergency Department Visit is represented with the QDM datatype and value set of Encounter, Performed: Emergency Department Visit (OID: 2.16.840.1.113883.3.117.1.7.1.292)

LDL-c

- LDL-c is represented with the QDM datatype and value set of Laboratory Test, Performed: LDL-c (OID: 2.16.840.1.113883.3.117.1.7.1.215)

Denominator Exceptions Data Elements:

Reasons for Not Prescribing Statin Medication

- Statin Allergy is represented with the QDM datatype and value set of Medication, Allergy: Statin Allergen (OID: 2.16.840.1.113883.3.117.1.7.1.423)
- Statin Ingredient Specific Medication is represented with the QDM datatype and value set of Medication, Discharge: Statin ingredient specific (OID: 2.16.840.1.113762.1.4.1021.7)
- Medical Reason is represented with the QDM datatype and value set of Medication, Discharge not done: Medical Reason (OID: 2.16.840.1.113883.3.117.1.7.1.473)
- Patient Refusal is represented with the QDM datatype and value set of Medication, Discharge not done: Patient Refusal (OID: 2.16.840.1.113883.3.117.1.7.1.93)

Non-Elective Inpatient Encounter

- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable, the measure is not stratified.

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable.

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
Not applicable.

S.16. Type of score:
Rate/proportion
If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
See attached HQMF file.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
eMeasures are calculated using only the structured data collected in certified EHR technology (CEHRT). Data not present in the structured field from which the measure draws will not be included in the measure calculation.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Imaging/Diagnostic Study, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Pharmacy

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Hospitals report EHR data using Certified Electronic Health Record Technology (CEHRT), and by submitting Quality Reporting Document Architecture Category 1 (QRDA-1).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF0439_CMS105v3_STK6_Bonnie_Testing.xlsx,STK6_eCQM_testing_attachment.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2836
Measure Title: STK-06: Discharged on Statin Medication
Date of Submission: 1/14/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	X Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the sub criteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are

present at start of care; [14,15] and has demonstrated adequate discrimination and calibration
OR
- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**

☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒other: Bonnie Test Cases	☒ other: Bonnie Test Cases (see question 1.2)

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

This measure is a legacy eCQM (an NQF endorsed measure that has been re-specified into an eMeasure and is currently used in a federal quality program/programs). In accordance with the modified NQF requirements for testing of eMeasures, the Bonnie testing tool was used to simulate a testing environment where measure specifications and HQMF output are tested against synthetic test data. Measure developers rely on the results in Bonnie to confirm whether the measure logic is performing as expected. Reference the eCQI Resource Center website (https://ecqi.healthit.gov/ecqm-tools/tool-library/bonnie) or the Bonnie testing tool website (https://bonnie.healthit.gov/) for more information about Bonnie functionality and its role in measure development. Please also reference the Bonnie testing worksheet attachment for detailed Bonnie test cases and testing results for this measure.

While Bonnie does consume HQMF specifications, we are not able to test the measure with data from HQMF implemented in EHRs.

1.3. What are the dates of the data used in testing? The measure specifications were tested in the Bonnie testing environment which mimics the year 2012.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☒ other: Bonnie Test Cases

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Not applicable.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

28 unique synthetic patient records were created in the BONNIE testing system for this measure. Cases were used to test the validity of each data element and timing relationship in the measure. Patient characteristics such as age, diagnosis, and lab results were pre-determined to provide a variety of scenarios that adequately tested for patients passing each data element and failing each data element. Data included in cases and tested for this measure included

diagnoses, medications, laboratory results, discharge statuses, patient orders, age, length of stay, and ED and inpatient encounters.

All 28 cases passed or failed as expected based on the data included in the case, confirming the measure logic is accurate and valid.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Not applicable. Bonnie test patients were developed for use across the different aspects of testing.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Not applicable.

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*May be one or both levels*)

X **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

The chart-abstracted version of this measure has been in national use since the 4th quarter of 2009. At the time the chart-abstracted measure was originally tested, extensive tests of measure reliability were conducted. At present, no performance data for the electronic version of the measure are currently available. Below is the reliability testing summary for NQF #0439 STK-06: Discharged on Statin Medication, from which this measure is derived.

At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with data transmission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of patient records over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rate of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify potential measure enhancements.

This measure was re-tested in 2010-2011 (n=33), following modification of the measure to focus specifically on statin medications (as opposed to the more general designation of lipid lowering medications, as had been originally tested). Retrospective comparison of vendor re-abstracted data and data submitted to The Joint Commission warehouse by the healthcare organization revealed zero defects. The agreement rate was 100% for all STK-6 Discharged on Statin Medication data elements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures). The following is a list of the major tests done on the submitted ORYX data:

- Transmission of complete data
- Usage of data received: To understand if the HCO provides the relevant service to treat the relevant population
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity
- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors as outlined in the contract is as follows:

- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data element agreement rates reported to The Joint Commission for the chart-abstracted version of this measure for the time period of one year (4Q2010 – 3Q2011) have shown an overall agreement rate of 96.7%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-6.

Data Elements	Total 'n' numerator	Total 'n' denominator	Rate
Clinical Trial	712	714	99.7%
Comfort Measures Only	709	714	99.3%
Elective Carotid Intervention	711	714	99.6%
LDL-c Greater Than or Equal to 100 mg/dL	363	380	95.5%
LDL-c Measured Within the First 48 Hours or 30 Days Prior to Hospital Arrival	434	463	93.7%
Pre-Arrival Lipid-Lowering Agent	426	464	91.8%
Reason for Not Prescribing Statin Medication at Discharge	91	99	92.0%
Statin Medication Prescribed at Discharge	416	441	94.3%

These agreement rates are considered to be well within acceptable levels.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i*.e., what do the results mean and what are the norms for the test conducted?*)

Agreement rates for individual data elements tested for the chart-abstracted version of this measure were within acceptable levels. Once data are available for analysis, it is expected that reliability tests of the eCQM version of this measure will yield similar results.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

- ☒ **Critical data elements** (*data element validity must address ALL critical data elements*)
- ☒ **Performance measure score**
  - ☐ **Empirical validity testing**
  - ☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The Bonnie testing tool and environment were used to test the measure logic and value sets. Each data element and logic statement was tested to confirm actual results met expectations. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure. An example of negative testing would be to include test cases with pediatric ages to ensure that pediatric patients are not included in the measure.

Denominator test cases positively test to ensure patients with principal diagnosis of ischemic stroke pass and fall into the measure. We have created negative test cases, testing to ensure patients without a principal diagnosis of ischemic stroke do not fall into the denominator.

Numerator test cases positively test to ensure patients receiving statin medication at discharge fall in to the measure. Negative test cases ensure that a patient who does not receive statin at discharge do not fall in to the numerator population.

Denominator exclusion and exception test cases for this measure ensure that patients are properly removed from the denominator if they have comfort measures, specific discharge statuses other than discharged home, and medical reasons or patient refusal for not receiving statin therapy. Negative test cases are also run. For example, cases with comfort measures ordered but not within the specified time frame are expected to fail, and should not be removed from the denominator. Testing confirmed patients meeting the exclusion and exception criteria are removed from the measure appropriately, while those that do not meet the criteria are retained in the denominator population.

A review of the measure specifications was also conducted to confirm the logic was properly expressed within the current version of the QDM and confirmed the logic matches the clinical intent of the measure, as stated in the measure header. This is done through use of a logic checklist to facilitate review of the measure logic, according to several checks, a few examples of which are included below:

- Is the intent of the measure described in the measure description articulated/ captured in the measure logic?
- Do the logic elements map to definitions in the measure narrative, data dictionary or supporting reference documentation?
- Do the populations in the narrative align with the populations defined in the logic?
- Are the mathematic inequalities reflective of the measure intent and represent the intended populations (for example: when intended the inequality represents less than rather than less and or equal to)?

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Bonnie results provide coverage and passing rates. This measure reached 100% coverage, confirming there is a test case for each pathway of logic. This measure also has a 100% passing rate, confirming all test cases performed as expected.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

This measure performed as expected in Bonnie.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Exclusions in the eCQM align with the chart-based version of the measure, and are clinically necessary for the interpretation of the measure. As noted previously, the chart-abstracted version of this measure has been in national use since the 4th quarter of 2009, and no data are available for the eCQM version of the measure. The below analysis addresses exclusions testing performed for the chart-abstracted version of the measure from which this measure is derived.

These exclusions were analyzed for frequency of occurrence. An issue that is of great concern to users of this measure is that due to the presence of exceptions to the measure, attainment of a 100% measure rate is not possible. Because of the role of this measure in the current Joint Commission accreditation process this is especially troubling to measure users. This concern is the basis for a number of the non-evidence-based exclusions to these measures. The following measure exclusions that were not derived directly from the evidence are as follows:

1. Patients < 18 years old
2. Patients who have a length of stay (LOS) greater than 120 days
3. Patients with Comfort Measures Only documented
4. Patients enrolled in clinical trials
5. Patients admitted for Elective Carotid Intervention
6. Patients discharged to another hospital (acute care facility)
7. Patients who left against medical advice
8. Patients who expired
9. Patients discharged to home for hospice care
10. Patients discharged to a health care facility for hospice care
11. Patients with a documented Reason For Not Prescribing Statin Medication at Discharge

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are given below.

Exclusion: Comfort Measures Only
Overall Occurrence n = 163,225
Overall Occurrence Percentage: 10.4%
Minimum: 0.30%
10th Percentile: 5.26%
Median: 10%
90th Percentile: 15.8%
Maximum: 35.7%

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%
10th Percentile: 0.24%
Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for Elective Carotid Intervention
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.5%
90th Percentile: 25.3%
Maximum: 95.2%

Exclusion: Patients with a documented Reason For Not Prescribing Statin Medication at Discharge
Overall Occurrence n = 13,297
Overall Occurrence Percentage: 4.2%
Minimum: 0.24%
10th Percentile: 1.11%
Median: 3.31%
90th Percentile: 10.5%
Maximum: 29.7%

Exclusion: Discharge Disposition - Patients discharged to another hospital
Overall Occurrence n = 449,924
Overall Occurrence Percentage: 35.7%
Minimum: 0.787%
10th Percentile: 25%
Median: 35.4%
90th Percentile: 46%
Maximum: 76.2%

Exclusion: Discharge Disposition - Patients who left against medical advice
Overall Occurrence n = 8,396
Overall Occurrence Percentage: 0.67%
Minimum: 0.067%
10th Percentile: 0.34%
Median: 0.86%

90th Percentile: 2.4%
Maximum: 9.67%

Exclusion: Discharge Disposition - Patients who expired
Overall Occurrence n = 76,168
Overall Occurrence Percentage: 6.04%
Minimum: 0.398%
10th Percentile: 1.9%
Median: 5.08%
90th Percentile: 10.2%
Maximum: 20.1%

Exclusion: Discharge Disposition - Patients discharged to home for hospice care
Overall Occurrence n = 658,264
Overall Occurrence Percentage: 52.2%
Minimum: 6.25%
10th Percentile: 39%
Median: 51.9%
90th Percentile: 64%
Maximum 94.3%

Exclusion: Discharge Disposition - Patients discharged to a health care facility for hospice care
Overall Occurrence n = 37,804
Overall Occurrence Percentage: 3%
Minimum: 0.169%
10th Percentile: 0.87%
Median: 3.01%
90th Percentile: 6.4%
Maximum: 23%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Analysis of these data for the chart-asbtracted progenitor of this measure indicated that all exclusions were appropriate. It is believed that results for exclusions in the eCQM will be similar when sufficient data have been received to perform such an analysis.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

- ☒ **No risk adjustment or stratification**
- ☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
- ☐ **Stratification by** Click here to enter number of categories **risk categories**
- ☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence

interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

A similar methodology will be used for the eCQMs.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

NQF#0439: STK-6 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10th Percentile
3Q2011	152	0.92074	0.14322	1	1	0.96214	0.9	0.82353
2Q2011	155	0.90883	0.16942	1	1	0.95833	0.9125	0.77778
1Q2011	159	0.87312	0.18575	1	1	0.9375	0.84615	0.6
4Q2010	136	0.87766	0.17491	1	0.98626	0.93333	0.83667	0.69231
3Q2010	126	0.89971	0.17615	1	1	0.96	0.9	0.7
2Q2010	117	0.90979	0.15111	1	1	0.96667	0.88889	0.72727
1Q2010	97	0.9034	0.15012	1	1	0.95652	0.875	0.67273
4Q2009	48	0.88567	0.18656	1	1	0.94949	0.85165	0.66667

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

It should be noted that since data collection on this measure is completely voluntary for both The Joint Commission and PCNASR, the hospitals reporting on this measure are self-selected, and therefore, presumably have a particular interest and concern for improving stroke care. For this reason, the measure results demonstrated by this group most likely significantly overstates the rate for the population of all health care organizations.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
***If only one set of specifications, this section can be skipped*.**

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Multiple data sources are not used for this measure.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Data not present in the structured field from which the measure draws will not be included in the measure calculation. In the Bonnie testing environment, missing data are tested as an expected "Fail" for that data element, and actual performance (whether or not the case fails) is compared to expectations to ensure missing data impacts data element and overall measure calculation as expected. This is the extent of missing data analysis that can be performed in Bonnie.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

See response for question 2b7.1 above.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

The measure performed as expected in the Bonnie testing environment.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment **Attachment:** STK6_eCQM_NQF_Measure_Feasibility_Assessment_Report.docx

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Not applicable.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
Value sets are housed in the Value Set Authority Center (VSAC), which is provided by the National Library of Medicine (NLM), in coordination with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

Viewing or downloading value sets requires a free Unified Medical Language System® (UMLS) Metathesaurus License, due to usage restrictions on some of the codes included in the value sets. Individuals interested in accessing value set content can request a UMLS license at (https://uts.nlm.nih.gov/license.html)

There are no other fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program http://jointcommission.org EHR Incentive Program https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

- Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
- Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 hospitals (2015)

- Name of program and sponsor: EHR Incentive Program; Centers for Medicare & Medicaid Services
- Purpose: The American Recovery and Reinvestment Act of 2009 (ARRA) (Pub.L. 111–5) was enacted on February 17, 2009. Title IV of Division B of ARRA amends Titles XVIII and XIX of the Social Security Act (the Act) by establishing incentive payments to eligible professionals (EPs), eligible hospitals, and critical access hospitals (CAHs), and Medicare Advantage Organizations to promote the adoption and meaningful use of interoperable health information technology (HIT) and qualified electronic health records (EHRs). These incentive payments are part of a broader effort under the HITECH Act to accelerate the adoption of HIT and utilization of qualified EHRs.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4,847 active registrations (September, 2015)

- Name of program and sponsor Hospital Accreditation Program; The Joint Commission
- Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
- Geographic area and number and percentage of accountable entities and patients included Nationwide; 3300 Joint Commission-accredited hospitals (2014)

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not applicable.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance

results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

While the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are currently available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs, one of which is this measure. Data will be accepted through February 28th, 2017.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Not applicable.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0074 : Chronic Stable Coronary Artery Disease: Lipid Control
0118 : Anti-Lipid Treatment Discharge
0439 : STK-06: Discharged on Statin Medication
0543 : Adherence to Statin Therapy for Individuals with Cardiovascular Disease
0545 : Adherence to Statins for Individuals with Diabetes Mellitus
0547 : Diabetes and Medication Possession Ratio for Statin Therapy
0639 : Statin Prescribed at Discharge
1519 : Statin Therapy at Discharge after Lower Extremity Bypass (LEB)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0543 : Adherence to Statin Therapy for Individuals with Cardiovascular Disease; Centers for Medicare and Medicaid Services
0545 : Adherence to Statins for Individuals with Diabetes Mellitus – measure not in NQF database
0547 : Diabetes and Medication Possession Ratio for Statin Therapy; CMS
0639 : Statin Prescribed at Discharge; CMS

5a. Harmonization

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**

Are the measure specifications completely harmonized?

No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

Three statin therapy measures were identified from the NQF database. All three measures address target diagnoses other than ischemic stroke or specific surgical procedures for patients 18 years or older: 0074 Coronary Artery Disease; 0118 isolated Coronary Artery Bypass Graft (CABG); and, 1519 Lower Extremity Bypass (LEB). Measure 1519 addresses inpatient organizational performance. The other two measures, 0074 and 0118 are provider-level measures in the ambulatory care setting. NQF# STK-06: Discharged on Statin Medication: The measures are completely harmonized to the extent possible, given the fact that the data source for #0439 is the paper medical record, and the data source for #2836 is the electronic health record.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Not applicable.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2837
Measure Title: STK-10: Assessed for Rehabilitation
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of ischemic or hemorrhagic stroke patients assessed for or who received rehabilitation services during the hospital stay. This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-4: Thrombolytic Therapy, STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-6 Discharged on Statin Medication, and STK-8: Stroke Education) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs.
Developer Rationale: Stroke is a leading cause of serious, long-term disability, associated with significant costs. The primary goal of rehabilitation is to prevent complications, minimize impairments, and maximize function. Evidence suggests that better clinical outcomes are achieved when post-acute stroke patients receive coordinated, multidisciplinary evaluation and intervention through the provision of rehabilitation services. Therefore, the need for rehabilitation services after an ischemic or hemorrhagic stroke should be assessed prior to discharge from the acute care setting.

Healthcare organizations that track rehabilitation assessments for internal quality improvement purposes have seen an improvement in the measure rate over time. The chart-asbtracted measure has been included in the Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvement in quality at the national level.

Numerator Statement: Ischemic or hemorrhagic stroke patients assessed for or who received rehabilitation services.
S.7. **Denominator Statement:** Patients age 18 and older discharged from inpatient care (non-elective admissions) with a principal diagnosis of ischemic or hemorrhagic stroke and a length of stay less or equal to 120 days.
S.10. **Denominator Exclusions:** Patients with comfort measures documented
Patients admitted for elective carotid intervention. This exclusion is implicitly modeled by only including non-elective hospitalizations.
Patients discharged to another hospital
Patients who left against medical advice
Patients who expired
Patients discharged to home for hospice care
Patients discharged to a health care facility for hospice care
Denominator Statement:
Denominator Exclusions:

Measure Type: Process
S.23. **Data Source:** Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
S.26. **Level of Analysis:** Facility, Population : National
Data Source:
Level of Analysis:

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

Evidence for this new eMeasure is the same as the existing measure #0441 STK 10: Assessed for Rehabilitation:

- Evidence suggests that better clinical outcomes are achieved when post-acute stroke patients receive coordinated, multidisciplinary evaluation and intervention through the provision of rehabilitation services. Therefore, the need for rehabilitation services after an ischemic or hemorrhagic stroke should be assessed prior to discharge from the acute care setting.
- AHA/ASA-Endorsed Management of Adult Stroke Rehabilitation Care: A Clinical Practice Guideline (2005).
  - Level I Evidence, Recommendation A - Organized and coordinated post-acute inpatient rehabilitation care improves outcome (page e-120).
  - Strongly recommend that once the patient is medically stable, the primary physician consult rehabilitation services (i.e., physical therapy, occupational therapy, speech and language pathology, kinesiotherapy, and physical medicine), as indicated, to assess the patient's rehabilitation needs and to recommend the most appropriate setting to meet those needs (page e-119).
  - The 2012 Committee expressed no concerns regarding the evidence underlying this measure.

Exception to evidence

Not applicable

Questions for the Committee:

- *The developer attests the underlying evidence for the measure #0441 (chart-abstracted version) has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the chart-abstracted measure has not changed and can be applied to the eMeasure, therefore there is no need for repeat discussion and voting on Evidence?*

Guidance from the Evidence Algorithm: Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: Moderate; Consistency: Moderate/High (Box 5) → Box 5b Moderate

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer states no performance data for the electronic version of this measure are yet available.
- This eCQM is currently included in the Hospital Inpatient Quality Reporting Program (IQR) and EHR Incentive program.
  - In CY2016, CMS is requiring organizations participating in HIQR to electronically submit one quarter of

data on 4 of 28 available eCQMs. This measure is one of the 28 eCQMs.

- The developer also provided performance data from the chart abstracted measure which seems to be very high with little to no room for improvement.

Disparities

- The developer does not provide disparities data from use of this measure.
- Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age."

Questions for the Committee:

- *Is there a gap in care that warrants a national performance eMeasure?*
- *How can this measure be used to better understand disparities in care and outcomes for certain groups?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **the measure is consistent with a number of stroke guidelines and is supported the body of evidence reflects upon studies that focused on coordinating or stroke ward care. Quality of evidence is moderate at max. Performance metric is the logical category

**this is a process measure that is well captured. It captures the desired factor - eval for rehabilitation services

**Evidence supports Emeasure

**The evidence basis for the chart-abstracted measure has not changed and can be applied to the eMeasure.

**No vote - same as 0441.

1b. Performance Gap

Comments: **performance data has been measured and appears achievable-- perhaps however at a ceiling effect. CMS is requiring this metric

Prior measures have shown this measure to be robust from a compliance perspective

**yes performance measures shown has grown and hospital performance improved. One concern is the ceiling effect that we may reach

**No performance data on this emeasure but high participation on chart abstracted data.

**There is a gap in care that warrants a national performance eMeasure. Many studies are cited that illustrate disparities in access to and utilization of rehabilitation services based on demographic characteristics, including age and race. This measure could be used to micro-segment the stroke population by demographic characteristics to illustrate opportunities to mitigate disparities in access to and utilization of rehabilitation services following stroke.

**Very high performance from chart version - little opportunity to improve.

There are disparities in care stroke and outcome, but no data presented regarding impact of assessment for rehabilitation prior to discharge.

1c. High Priority (previously referred to as High Impact)

Comments: **I do not think weighted rules are in place however, such a factor would be helpful if function becomes one of the weighted factors

**The stated metric is logical-- we are not as clear as to how it would fit into the overall 8 factor construct however it appears to be linked with 5 other performance metrics linked to better outcome

**NA

**N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Electronic Clinical Data, Electronic Health Record (EHR), Imaging/Diagnostic Study, Laboratory Data, and Pharmacy Data. This is an eMeasure.

Specifications:

- HQMF specifications for eMeasure are included in the document set on SharePoint. See eMeasure Technical Review below.
- The developer states that changes have been made to the eCQM specifications to reflect revisions to the chart abstracted measure from which this measure is derived (NQF measure #0441).

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor review:

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	N/A – All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM;
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously Demonstrated through Bonnie testing
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible. This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment.

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high

proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☒ **No**

Method(s) of reliability testing:

- Dataset used for testing included 24 synthetic records created in the BONNIE testing system. Measure not tested in EHR.
- Data element validity testing was performed and will count for data element reliability as well – see validity testing section.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results from the Bonnie tool demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm: Precise specifications (Box 1) →empiric reliability testing as specified (Box 2) → empiric validity testing of patient-level data (Box 3) → YES: Use rating from validity testing of patient-level data elements (Box 10)→Assessed the measure logic only (Box 11)→Low certainty or confidence data used in measure are valid (Box 12b)→Low (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Preliminary rating for reliability: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- The Bonnie testing tool and environment with 24 synthetic patient records were used to test the measure logic and value sets. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure.

- The numerator, denominator, exclusions/exceptions, and logic statement were tested to confirm actual results met expectations.
- The developer also conducted a review of the measure specifications to confirm the measure logic was properly expressed with the current version of the QDM and that the logic matches the clinical intent of the measure.

Validity testing results:

- The testing results from the Bonnie tool reached 100% coverage and confirmed there was a test case for each pathway of logic (negative and positive test cases).
- The measure also had a 100% passing rate which confirmed that all the test cases performed as expected.
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer submitted additional information on exclusions and meaningful differences.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Is it reasonable to assume that the impact of the exclusions will be similar for the eMeasure and the chart-abstracted measure?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer submitted additional information on exclusions and meaningful differences.

Question for the Committee:

- *For meaningful differences in the chart-abstracted measure the developer refers to the performance scores over time. Does the SC think the eMeasure will demonstrate similar results to the chart-abstracted measure which show little to no room for improvement?*

2b6. Comparability of data sources/methods: Not applicable

2b7. Missing Data

- The developer describes the handling of missing data but does not provide information on the frequency of missing data or potential impact on measure results due to limited testing abilities using the Bonnie tool.

Guidance from Validity Algorithm: Specifications consistent with evidence (Box 1) → threats to validity empirically assessed to some extent (Box 2) → empirical testing of data elements using BONNIE tool →Moderate

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **the measure appears to be consistent with the evidence

**I think it is reasonable from a validity perspective and we note that the exclusions are reasonable and reflect factors

**No concerns

**The specifications are consistent with the evidence.

**Data elements here are clear as well.

2a2. Reliability Testing

Comments: **The validity is established with this subset metrics

**Validity testing suggests a relationship with 6 other stroke metrics that are associated with high quality

**Data element testing done, no face validity testing done

**The sample of 24 is not adequate to generalize for widespread implementation. A larger sample from the Bonnie tool would demonstrate greater validity.

**24 synthetic tests - similar to chart data. No face validity included. While the results of these tests are positive the numbers are smaller than would have been desired. In the end the question is whether the assessments actually results in the prompt and proper rehabilitation and whether that actually improves the recovery and quality of life.

2b2. Validity Testing

Comments: **No

**Missing data is small and the exclusions are logical and do not appear to be a threat to the validity

**Threats to validity were not addressed

**Missing data is small and not identified as a threat to validity.

**One could easily look for much more data - including disparity data. The limited data provide many varieties of data that might be of interest and identify additional subtleties in the relationship of assessment and rehabilitation.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **I need further clarity in this area ie what is expected

**Inter rater between hospitals appears to be reliability is 98.3%

**Low reliability

**The sample of 24 is not adequate to generalize for widespread implementation. A larger sample from the Bonnie tool would demonstrate greater reliability.

**No data regarding consistency of application. Are there different assessments, are they performed with consistency?

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is EHR; all data elements are in defined fields in EHR.
- The submission contains a feasibility assessment that addresses data element feasibility and follow-up with the measure developer indicates that the measure logic is feasible.
- This is a legacy eMeasure included in the Meaningful Use program. The developer submitted Bonnie testing results to establish feasibility, and provided an interpretation of the testing process and results in the measure testing attachment.
- Per developer, they were unable to directly assess data accuracy of the data elements in an EHR system. Workflow, or the degree to which the data elements are captured during the course of care was also not assessed.

Questions for the Committee:

- *Is it likely the required data elements are routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **the measure is very feasible to collect as evidenced by compliance data

**the national rate of reporting is 98.7% thus the metric is feasible

**Acceptable

**All data elements are available in the EHR. It is likely that the required data elements are routinely generated and used during care delivery. The data collection strategy is ready to be put into operational use. The eMeasure Feasibility Score Care demonstrates acceptable feasibility in multiple EHR systems and sites.

**Since the data should all be available electronically this should be very feasible. Quality of data in the EHRs is always a questions, as is compatibility of various EHR systems.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- CMS Hospital Inpatient Quality Reporting Program (IQR)/EHR Incentive Program: The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates. The EHR Incentive Program provides incentive payments to promote the adoption and meaningful use of interoperable HIT and qualified EHRs.
- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results:

- The developer states that while the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are currently available.
- In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs, one of which is this measure. Data will be accepted through February 28th, 2017.

Unexpected findings (positive or negative) during implementation:

- Developer did not identify unexpected findings during implementation.

Potential harms:

- Developer did not identify any unintended consequences related to this measure.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

○ *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **the measure is reported in an identifiable and usable manner

**the metric is reported in CMS compare, and in various stroke registers as well as the JCAHO annual report

**No performance data available; benefits outweigh unintended consequences

**The performance results can be used to further the goal of high-quality, efficient healthcare by measuring conformance to this process measure that has evidence to support improved outcomes for stroke patients. No unintended consequences are identified.

**Data will be publicly reported - the real use of the data will be the demonstration of the value of the assessment and subsequent rehabilitation on the patient recovery and quality of life.

Criterion 5: Related and Competing Measures

Related or competing measures:

- 0244 : Stroke and Stroke Rehabilitation: Rehabilitation Services Ordered
- 0441 : STK-10: Assessed for Rehabilitation

Harmonization:

- The developer states NQF#0244 focuses on rehabilitation orders written prior to hospital discharge and not the rehabilitation assessment or services received by the patient. NQF#0441:STK-10 Assessed for Rehabilitation. The measures are completely harmonized to the extent possible, given the fact that the data source for #0441 is the paper medical record, and the data source for #2837 is the electronic health record.

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0441 **NQF Project:** Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.
(evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The focus of the measure is to prevent complications, minimize impairments, and maximize function post stroke by assessing the patient's need for rehabilitation services and/or initiating rehabilitation.
Assessed for rehabilitation need >> rehabilitation services received >> improved neurological outcomes >> decreased morbidity and mortality.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
This measure is consistent with the adult stroke rehabilitation care guidelines developed for the Veterans Health Administration and Department of Defense Medical Systems and endorsed by the American Heart Association/American Stroke Association (2005). Guideline recommendations incorporated information from existing evidence-based guidelines from the Agency for Health Care Policy and Research (AHCPR) Post-Stroke Rehabilitation (1995), the Scottish Intercollegiate Guidelines Network (SIGN) Management of Patients with Stroke (1997), and the Royal College of Physicians (RCP) National Clinical Guidelines for Stroke (2000).

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: According to the 2011 update from the Evidence-Based Review of Stroke Rehabilitation (EBRSR), 2000 studies including 1, 078 randomized control trials (RCTs) have been identified in the stroke rehabilitation literature. The 2001 EBSRS identified five meta-analyses of the effectiveness of stroke rehabilitation:

- Langhorne, et al. (1993) – 10 RCTs conducted between 1962 and 1993. Results revealed that management of stroke patients on a stroke unit was associated with lower mortality rates than general medical wards, and a 28% reduction in the risk of death occurring in the first 17 weeks post-stroke.
- Ottenbacher and Jannell (1993) – 36 RCTs (n=3,717 patients). Patients who participated in an individualized stroke rehabilitation programs performed better than 65% of those patients in the comparison group. Greater functional improvements were observed in younger patients and those with relatively short stroke onset to rehabilitation admission intervals.
- Stroke Unit Trialists' Collaboration (2007) - 31 RCTs. Stroke unit care was associated with a significant reduction in death (OR 0.86; 95% CI 0.76-0.98; P=0.02) at a median of one-year follow-up. Stroke unit care was also associated with a significant reduction in the combined outcomes or both death or institutional care (OR 0.82; 95% CI 0.73-0.92; P=0.0006) and death or dependency (OR 0.82; CI 95% 0.73-0.92, P=0.001).
- The Canadian Coordinating Office of Health Technology Assessment (CCOHTA) (2003) – 6 RCTs (n=1,709 patients) from 1995 to July 2002. Stroke unit care was associated with a reduction in the odds of death (OR 0.60; CI 95% 0.42-0.86) an outcome recorded in all studies.
- Seenan, et al. (2007) – 18 non-randomized trials, which more closely approximate usual clinical practice. Findings reported

were similar to prior meta-analyses of RCTs. The odds of death (OR 0.79; 95% CI 0.72-0.86) and poor outcome (OR 0.87; CI 9%%0.80-0.95) were reduced for patients receiving stroke unit care compared to general medical management.

In addition, a total of 37 individual studies (14 non-randomized and 23 RCTs) of the efficacy of stroke rehabilitation were included in the review. Only the results of RCTs and quasi RCTs were used to formulate conclusions. Studies were categorized according to the type of care provided ranging from acute stroke management to settings that offered various levels of rehabilitation services (i.e., combined acute and rehabilitation stroke units, subacute rehabilitation units, and mobile stroke teams). Overall, the body of evidence strongly supports that units providing stroke rehabilitation are associated with improved functional outcomes for patients.

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting stroke rehabilitation is moderate. The impact of stroke rehabilitation on outcomes has been difficult to quantify due to problems with study design and methodology (e.g., lack of randomization, inappropriate control group selection, failure to blind assessors, difficulty in controlling for all possible confounders) detected by systematic review. Furthermore, issues inherent to stroke rehabilitation, such as controlling for spontaneous neurological recovery, daily fluctuation in individual function, and difficulties measuring functional outcomes have challenged study designs. Pre-selection of patients and observer measurement bias are additional concerns when studying the impact of stroke rehabilitation on outcomes (Foley, et al, 2011).

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of evidence consistently supports that rehabilitation services minimize functional impairment and improve clinical outcomes for patients following stroke. Furthermore, most stroke patients have some residual deficits following the event, and can benefit from rehabilitation services tailored to improved cognitive, speech, or motor function. Based on these findings, a rehabilitation assessment should be completed for all stroke patients prior to hospital discharge.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)***:**
There is strong evidence on the net benefits of stroke rehabilitation. The body of evidence clearly supports that stroke rehabilitation is associated with a reduction in the odds of death or dependency and the need for institutionalization. There is strong evidence that stroke rehabilitation improves functional outcomes. There is also strong evidence that stroke rehabilitation reduces mortality for a subset of more severe stroke patients.

There is no dispute that stroke represents a significant economic burden in developed countries; however, there is some uncertainty respecting the cost-effectiveness of stroke rehabilitation. Most of the studies have been conducted outside of the United States in the United Kingdom, Australia, or the Netherlands. Stroke recovery and residual disability are highly variable. Studies have ignored the contributions of family members and/or caregivers towards rehabilitative progress, and the discrete components of caregiver and associated costs are difficult to isolate.

Although findings differ, there is some evidence of the cost-effectiveness of stroke rehabilitation services. Kalra et al. (2005) found stroke unit care to be more effective than home care, and also to be of equal cost (using per patient day alive), suggesting that stroke unit care is more cost-effective than home care. Using prospectively collected data from the Stroke Care Outcome: Providing Effective Services (SCOPES) trial over six months, Moodie et al. (2007 - Australia) compared the effectiveness of stroke units, conventional care, and a mobile service. While better outcomes were noted for patients cared for in stroke units, the incremental costs were higher compared to conventional care. Saka et al. (2009 – UK) projected the cost-effectiveness of three types of care over a 10-year period: stroke units with early supported discharge (ESD), stroke units without ESD, and general medical-surgical care. Although the costs of care were greater for both stroke units when compared to general medical-surgical wards, the cost per quality-adjusted life years (QALY) was lowest for stroke units with ESD (Foley, et al., 2011).

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: The Evidence-Based Review of Stroke Rehabilitation (EBRSR) funded by the Canadian Stroke Network (CSN).

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: Canadian Stroke Network; 2010 Dec 8. p. 129-150.
Rating Scheme for the Strength of the Evidence
Summary of Definitions for Levels of Evidence*
Grade A – Criteria – Strong recommendation. Evidence from randomized controlled trials or meta-analysis of randomized controlled trials. Desirable effects clearly outweigh undesirable effects, or vice versa.
Grade B – Criteria – Single randomized controlled trial or well-designed cohort or case-control analytic study; or multiple time series or dramatic results of uncontrolled experiment. Desirable effects closely balanced with undesirable effects.
Grade C – Criteria – At least one well-designed, nonexperimental descriptive study (e.g., comparative studies, correlation studies, case studies) or expert committee reports, opinions and/or experience of respected authorities, including consensus from development and/or reviewer groups.
*Based on Guyatt GH, Cook DJ, Jaeschke R, et al. Grades of recommendation for antithrombotic agents: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th edition). [published erratum in Chest 2008;34:47]. Chest 2008;133(6 Suppl):123S-131S.

1c.13 Grade Assigned to the Body of Evidence: Grade A Strong Evidence

1c.14 Summary of Controversy/Contradictory Evidence: The benefits of stroke rehabilitation are obvious. No controversies over the benefits of stroke rehabilitation were noted in the literature.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- Foley N, Teasell R, Bhogal S, Speechley M. The efficacy of stroke rehabilitation. The Evidence-Based Review of Stroke Rehabilitation. August 2011: 1-50.
- Kalra L, Evans A, Perez I, Knapp M, Swift D, Donaldson N, Swift CG. Alternative strategies in stroke care. Health Technology Assessment. 2005;9:1-94.
- Keith RA. Rehabilitation after stroke: cost-effectiveness analyses. Journal of the Royal Society of Medicine. 1996;89:631-633.
- Langhorne P, Williams BO, Gilchrist W, Howie K. Do stroke units save lives? Lancet. 1993;342:395-398.
- Moodie M, Cadilhac D, Pearce D. Economic evaluation of Australian stroke services: a prospective, muticenter study comparing dedicated stroke units with other care modalities. Stroke. 2006:37:2790-2795.
- Noorani HZ, Brady B, McGahan L, Teasell R, Skidmore B, Doherty T. Stroke rehabilitation services: systematic reviews of the clinical and economic evidence. Canadian Coordinating Office for Health Technology Assessment. March 2003; Technology Report No. 35.
- Ottenbacher KJ, Jannell S. The results of clinical trials in stroke rehabilitation research. Arch Neurol. 1993;50:37-44.
- Saka O, Serra V, Samyshkin Y, McGuire A, Wolfe CC. Cost-effectiveness of stroke unit care followed by early supported discharge, Stroke. 2009;40:24-49.
- Zorowitz RD, et al. the Post-Stroke Rehabilitation Outcomes Project (PSROP), Top Stroke Rehabil. 2005 Fall;12(4).

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)*:
AHA/ASA-Endorsed Management of Adult Stroke Rehabilitation Care: A Clinical Practice Guideline (2005).
Level I Evidence, Recommendation A
Organized and coordinated post-acute inpatient rehabilitation care improves outcome (page e-120).
1. Strongly recommend that once the patient is medically stable, the primary physician consult rehabilitation services (i.e., physical therapy, occupational therapy, speech and language pathology, kinesiotherapy, and physical medicine), as indicated, to assess the patient's rehabilitation needs and to recommend the most appropriate setting to meet those needs (page e-119).

1c.17 Clinical Practice Guideline Citation: Duncan PW, Zorowitz R, Bates B, Choi JY, Glasberg JJ, Graham GD, Katz RC, Lamberty K, Refer D. Management of Adult Stroke Rehabilitation Care: A Clinical Practice Guideline. Stroke. 2005;36:e100-143.

1c.18 National Guideline Clearinghouse or other URL: http://stroke.ahajournals.org/content/36/9/e100.full.pdf+html?sid=b27cb2e0-836f-4fc0-9ee0-aae14d24f3e4

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation

and any disclosures regarding bias: The Stroke Council of the American Heart Association has chosen to endorse the VA DoD guideline for stroke rehabilitation (Management of Stroke Rehabilitation. Washington, DC: VA/DoD Clinical Practice Guideline Workgroup, Veterans Health Administration, Department of Veterans Affairs and Health Affairs, Department of Defense, February 2003). The Department of Defense and Veterans Health Administration panel of experts evaluated the medical evidence according to criteria proposed by the US Preventive Services Task Force. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 **System Used for Grading the Strength of Guideline Recommendation:** USPSTF

1c.22 **If other, identify and describe the grading scale with definitions:**

1c.23 **Grade Assigned to the Recommendation:** A: A strong recommendation that the intervention is always indicated and acceptable.

1c.24 **Rationale for Using this Guideline Over Others:** The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: High 1c.26 Quality: Moderate1c.27 Consistency: Moderate

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

0441_Evidence_MSF5.0_Data-635827722695140517.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*

Stroke is a leading cause of serious, long-term disability, associated with significant costs. The primary goal of rehabilitation is to prevent complications, minimize impairments, and maximize function. Evidence suggests that better clinical outcomes are achieved when post-acute stroke patients receive coordinated, multidisciplinary evaluation and intervention through the provision of rehabilitation services. Therefore, the need for rehabilitation services after an ischemic or hemorrhagic stroke should be assessed prior to discharge from the acute care setting.

Healthcare organizations that track rehabilitation assessments for internal quality improvement purposes have seen an improvement in the measure rate over time. The chart-asbtracted measure has been included in the Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvement in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This measure is a legacy eCQM that is currently included in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data are yet available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs. This measure is one of the 28.

For more information, refer to the CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
In October, 2009, The Joint Commission added the chart-abstracted stroke (STK) measure set as a new core measure option to meet performance measure requirements for Joint Commission hospital accreditation purposes. Below is the specified level of analysis for STK-10 Assessed for Rehabilitation beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 2555 denominator cases; 2473 numerator cases; 49 hospitals; 0.96791 national aggregate rate; 0.95212 mean of hospital rates; 0.07659 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.97744 50th percentile rate/median rate; 0.95 25th percentile rate/lower quartile; and, 0.85294 10th percentile rate.

CY 2010: 23,875 denominator cases; 23,144 numerator cases; 138 hospitals; 0.96938 national aggregate rate; 0.95282 mean of hospital rates; 0.06331 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.97731 50th percentile rate/median rate; 0.93333 25th percentile rate/lower quartile; and, 0.875 10th percentile rate.

CY 2011: 28,415 denominator cases; 27,731 numerator cases; 157 hospitals; 0.97593 national aggregate rate; 0.96355 mean of hospital rates; 0.06802 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.98671 50th percentile rate/median rate; 0.96203 25th percentile rate/lower quartile; and, 0.89939 10th percentile rate.

CY 2012: 28,735 denominator cases; 28,272 numerator cases; 158 hospitals; 0.98389 national aggregate rate; 0.96919 mean of hospital rates; 0.06586 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.98917 50th percentile rate/median rate; 0.96875 25th percentile rate/lower quartile; and, 0.92857 10th percentile rate.

CY 2013: 44,242 denominator cases; 43,527 numerator cases; 262 hospitals; 0.98384 national aggregate rate; 0.97722 mean of hospital rates; 0.05526 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99131 50th percentile rate/median rate; 0.97605 25th percentile rate/lower quartile; and, 0.95652 10th percentile rate.

CY 2014: 210,075 denominator cases; 207,311 numerator cases; 1299 hospitals; 0.98684 national aggregate rate; 0.9772 mean of hospital rates; 0.06807 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99433 50th percentile rate/median rate; 0.98 25th percentile rate/lower quartile; and, 0.94737 10th percentile rate.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers,

immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al. 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, 2011).

In the national REGARDS cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages (Howard, et al., 2011).

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors. Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038) and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years if age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age: RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in
blacks or whites (Kleindorfer, et al., 2010).

In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).
Few studies have examined racial-ethnic disparities in stroke rehabilitation. Findings from the studies available are somewhat contradictory. With regard to access, a study from the Veterans Health Administration found no differences in referral to or receipt of inpatient rehabilitation (Horner RD, et al., 2003); however, another study found that urban-dwelling blacks or African Americans were more likely to be discharged from the hospital to inpatient rehabilitation facilities than whites (Gregory PC, et al, 2006). The VA study from Horner also reported that blacks or African Americans had a slightly longer time to initiation of rehabilitation that whites (4.4 days versus 3.8 days; $P<0.05$).

Other studies have evaluated the effect of the Functional Independence Measure (FIM) score on the admission of minorities to

inpatient rehabilitation facilities. No consistency of findings has been noted. A retrospective cohort study from Bhandari, et al. (2005) found no racial differences in admission FIM scores, but several other studies reported significant differences. One large study involving urban participants hospitalized at a county hospital reported that Hispanics had lower FIM scores than blacks or African Americans (Chiou-Tan FY, et al, 2006). Another study reported similar findings with whites having the highest FIM scores (Ottenbacher KJ, et al., 2008).

Rehabilitation length of stay has been studied with similarly inconsistent findings. One study of VA and non-VA inpatient rehabilitation facilities reported significantly longer lengths of stay for blacks or African Americans than for whites (Stineman MG, 2001). Another study from CDC (2007) using data from the BRFSS found that blacks or African Americans were more likely to be referred to outpatient rehabilitation services than whites (adjusted OR 1.49; 95% CI 1.1 to 2.0).

Results on the delivery of outpatient occupational and physical therapy are also conflicting with some reporting that whites are more likely to receive therapy than non-white minorities (Mayer-Oakes SA, 1992). Another study utilizing data from the Health and Retirement Study found no racial differences among patients receiving occupational or physical therapy (Cook C, 2005). Studies from Bhandari (2005), as well as Horner (1997), reported no racial differences in the intensity of occupational or physical therapy delivered to minorities when compared to whites.

A more recently published study (Qian F, et al, 2013) from the American Heart Association/American Stroke Association Get With The Guidelines (GWTG)-Stroke program from April 2003 through December 2008, (n=200,900), found racial disparities related to length of hospitalization and functional status. After multivariable adjustment for both patient-level and hospital-level characteristics, compared with non-Hispanic white patients, non-Hispanic black and Hispanic patients were more likely to have a longer hospital stay (i.e., length of stay greater than 4 days; black: AOR, 1.30; 95% CI, 1.22-1.39; P<0.001; Hispanic: AOR, 1.12; 95% CI, 1.02-1.23; P=0.02) and less likely to be ambulatory independent at discharge (black: AOR, 0.81; 95% CI, 0.76-0.87; P<0.001; Hispanic: AOR, 0.89; 95% CI, 0.79-0.99; P=0.04). Furthermore, non-Hispanic black patients were less likely to have a death at admission (black: AOR, 0.78; 95% CI, 0.68 to 0.89; P<0.001) or to be discharged home (black: AOR, 0.87; 95%CI, 0.81-0.94; P<0.001). Data from the Paul Coverdell National Acute Stroke Registry (PCNASR) found that other races were somewhat more likely to be assessed for rehabilitation services (98.9%) when compared to non-Hispanic white patients (98.1%) (Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, 2014).

• Bhandari VK, Kushel M, Price L, Schillinger D. Racial disparities in outcomes of inpatient stroke rehabilitation. Arch Phys Med Rehabil. 2005;86:2081-2086.
• Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, Yan T, Escarce JJ. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-3368.
• Centers for Disease Control and Prevention (CDC). Outpatient rehabilitation among stroke survivors: 21 staes and the District of Columbia, 2005. MMWR. 2007;56:504-507.
• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Chiou-Tan FY, Keng MJ Jr, Graves DE, Chan KT, Rintala DH. Racial/ethnic differences in FIM scores and length of stay for underinsured patients undergoing stroke inpatient rehabilitation. Am J Phys Med Rehabil. 2006;86:415-423.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Cook C, Stickley L, Ramey K, Knotts VJ. A variables associated with occupational and physical therapy stroke rehabilitation utilization and outcomes. J Allied health. 2005;34:3-10.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, Griffith P, Gorelick PB, Howard G, Leira EC, Morgenstern LB, Ovbiagele B, Peterson E, Rosamond W, Trimble B, Valderrama AL, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• Gillum RF, Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 2011;42(11):3294-3296.
• Gregory PC, Han E, Morozova O, Kuhlemeier KV. Do racial disparities exist in access to inpatient stroke rehabilitation in the state of Maryland? Am J Phys Med Rehabil.2006;85:814-819.
• Horner RD, Hoenig H, Sloane R, Rubenstein LV, Kahn KL. Racial differences in the utilization of inpatient rehabilitation services among elderly stroke patients. Stroke. 1997;28:19-25.
• Horner RD, Swanson JW, Bosworth HB, Matchar DB. VA Acute Stroke (VAST) Study Team. Effects of race and poverty on the process and outcome of inpatients rehabilitation services among stroke patients. Stroke. 2003;34:1027-1031.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-3375.

• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Jacobs BS, Birbeck G, Mullard AJ, Hickenbottom S, Kothari R, Roberts S, Reeves MJ. Quality of hospital care in African Americans and white patients with ischemic stroke and TIA. Neurology.2006;66:809-814.
• Johnston SC, Fung LH, Gillum LA, Smith WS, Brass LM, Lichtman JH, Brown AN. Utilization of intravenous tissue-type plasminogen activator for ischemic stroke at academic medical centers.: the influence of ethnicity. Stroke. 2001; 32:1061-1068.
• Karve S, Balkrishnan R, Seiber E, Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Mayer-Oakes SA, Hoenig H, Atchison KA, Lubben JE, De Jong F, Schweitzer SO. Patient-related predictors of rehabilitation use for community-dwelling older Americans. J AM Geriatr Soc. 1992;40:336-342.
• Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, Kerber KA, Skolarus LE, Meurer WJ, Burke JF, Adelman EE, Baek J, Lisabeth LD. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol. 2013;74:778–785.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Ottenbacher KJ, Campbell J, Kuo YF, Deutsch A, Ostir GV, Granger CV. Racial and ethnic differences in postacute rehabilitation outcomes after stroke in the United States. Stroke. 2008;39:1514-1519.
• Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Stineman MG, Ross RN, Hamilton BB, Maislin G, Bates B, Granger CV, Asch DA. Inpatient rehabilitation after stroke: a comparison of lengths of stay and outcomes in the Veterans Affairs and non Veterans Affairs health care system. Med Care. 2001;39:123-137.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, Patient/societal consequences of poor quality, Severity of illness

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States remains attributable to stroke. More women than men die of stroke each year. Women accounted for almost 60% of US stroke deaths in 2008 (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, et al., 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 (P<0.05).168 (US Burden of Disease Collaborators, 2013) . Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

The evidence indicates that patients do better with a well-organized, multidisciplinary approach to post-acute rehabilitation after stroke. The rehabilitation team may consist of a physician, nurse, physical therapist, occupational therapist, kinesiotherapist, speech and language pathologist, psychologist, recreational therapist, patient, and family members/caregivers. If an organized rehabilitation team is not available in the facility, the evidence indicates that patients with moderate or severe symptoms should be offered a referral to a facility with such a service.

Stroke rehabilitation should begin during the acute hospitalization, as soon as the diagnosis of stroke is established and life-threatening problems are controlled. The highest priorities of early stroke rehabilitation are to prevent recurrence of stroke, manage comorbidities, and prevent complications related to immobility, dysphagia, and bowel and bladder dysfunction. Rehabilitation services may include: dysphagia treatment and management; speech therapy for communication disorders (i.e., aphasia and dysarthria) and related cognitive impairments; lower-extremity strengthening and gait training; positioning, passive stretching, range-of-motion exercises, and pharmacotherapy for patients with paretic limbs and muscle spasticity; corrective measures (e.g., splinting, serial casting, surgery) for contractures; treatment interventions for post-stroke shoulder pain; treatment for depression and other cognitive and emotional disorders; and, other services. Recently published literature emphasizes the importance of identifying and managing cognitive-communication impairments after stroke since these impairments are associated with higher rate of death and higher risk of future stroke (Hinckley, 2014).

Living with disabilities after a stroke is a lifelong challenge during which people continue to seek and find ways to compensate for or adapt to persisting neurological deficits. For many, the real work of recovery begins after formal rehabilitation when the patient attempts to use newly learned skills without the support of the rehabilitation environment or team (Bates, 2005).

1c.4. Citations for data demonstrating high priority provided in 1a.3

• American Academy of Physical Medicine and Rehabilitation. Rehabilitation Helps Stroke Patients Recover Skills. AAPM&R Chicago, IL Office: Author.
• American Academy of Physical Medicine and Rehabilitation. Urgency Key But Perseverance Pays Off. AAPM&R Chicago, IL Office: Author.
• Bates B, Choi JY, Duncan PW, Glasberg JJ, Graham GD, Katz RC, Lamberty K, Recker D, Zorowitz R. American Heart Association/American Stroke Association-endorsed practice guideline. Veterans Affairs/Department of Defense clinical practice guideline for the management of adult stroke rehabilitation care. Stroke. 2005;36:2049.
• George M, Xin Tong, McGruder H, Yoon P, Rosamond W, Winquist A, Hinchey J, Wall H, Pandey D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Hinckley JJ. A case for the implantation of cognitive-communication screenings in acute stroke. American Journal of Speech-Language Pathology. 2014;23:4-14.
• Management of patients with stroke. Rehabilitation, prevention and management of complications, and discharge planning, Scottish Intercollegiate network Guidelines Network (SIGN), 2002.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• National Institute of Neurological Disorders. Post-Stroke Rehabilitation Fact Sheet. National Institute of Neurological Disorders Bethesda, MD Office: Author.
• Post Stroke Rehabilitation, Clinical Practice Guideline No.16, Agency for Health Care Policy and Research (now known as Agency for Healthcare Research and Quality), 1995.
• VA/DoD Clinical Practice Guideline for the Management of Stroke Rehabilitation in the Primary Care Setting, 2010.
• Zorowitz RD, et al. the Post-Stroke Rehabilitation Outcomes Project (PSROP), Top Stroke Rehabil. 2005 Fall;12(4).

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input*

was obtained.)
Not applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Functional Status

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/ecqm_library.html

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is an eMeasure **Attachment:** STK10_MAT.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** AssessedforRehabilitation_v4_Wed_Apr_01_13.26.48_CDT_2015-1-.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Changes have been made to the eCQM specifications in order to reflect revisions to the chart abstracted measure from which this measure is derived.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Ischemic or hemorrhagic stroke patients assessed for or who received rehabilitation services.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Rehabilitation Services is represented with any one of the following QDM datatypes and value sets:
- Procedure, Performed: Rehabilitation Assessment (OID: 2.16.840.1.113762.1.4.1045.18)
- Procedure, Performed: Rehabilitation Therapy (OID; 2.16.840.1.113762.1.4.1045.19)
Encounter, Performed: Non-Elective Inpatient Encounter (discharge status: Discharged to Rehabilitation Facility)
  o Non-Elective Inpatient Encounter SNOMEDCT Value Set (2.16.840.1.113883.3.117.1.7.1.424)

o Discharged to Rehabilitation Facility SNOMEDCT Value set (2.16.840.1.113883.3.117.1.7.1.132)

Patients who Refuse Rehabilitation Assessment
Patient Refusal is represented with the QDM datatype and value sets of Procedure, Performed not done: Patient Refusal for Rehabilitation Assessment
o Rehabilitation Assessment SNOMEDCT Value set (2.16.840.1.113762.1.4.1045.18)
o Patient Refusal SNOMEDCT Value set (2.16.840.1.113883.3.117.1.7.1.308)

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Patients age 18 and older discharged from inpatient care (non-elective admissions) with a principal diagnosis of ischemic or hemorrhagic stroke and a length of stay less or equal to 120 days.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Principal Diagnosis of Ischemic Stroke
- Ischemic Stroke is represented with the QDM datatype and value set of Diagnosis, Active: Ischemic Stroke (OID: 2.16.840.1.113883.3.117.1.7.1.247)
- Ordinality: Principal (OID: 2.16.840.1.113883.3.117.1.7.1.14)

Principal Diagnosis of Hemorrhagic Stroke
- Hemorrhagic Stroke is represented with the QDM datatype and value set of Diagnosis, Active: Hemorrhagic Stroke (OID: 2.16.840.1.113883.3.117.1.7.1.212)
- Ordinality: Principal (OID: 2.16.840.1.113883.3.117.1.7.1.14)

Non-Elective Inpatient Encounter
- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Patients with comfort measures documented
Patients admitted for elective carotid intervention. This exclusion is implicitly modeled by only including non-elective hospitalizations.
Patients discharged to another hospital
Patients who left against medical advice
Patients who expired
Patients discharged to home for hospice care
Patients discharged to a health care facility for hospice care

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Non-Elective Inpatient Encounter
- Non-Elective Inpatient Encounter is represented with the QDM datatype and value set of Encounter, Performed: Non-Elective Inpatient Encounter (OID: 2.16.840.1.113883.3.117.1.7.1.424)
Discharge Status (modeled as Attributes of the above Non-Elective Inpatient Encounter)
- Discharge status: Left Against Medical Advice (OID: 2.16.840.1.113883.3.117.1.7.1.308)
- Discharge status: Patient Expired (OID: 2.16.840.1.113883.3.117.1.7.1.309)
- Discharge status: Discharge To Acute Care Facility (OID: 2.16.840.1.113883.3.117.1.7.1.87)
- Discharge status: Discharged to Home for Hospice Care (OID: 2.16.840.1.113883.3.117.1.7.1.209)

- Discharge status: Discharged to Health Care Facility for Hospice Care (OID: 2.16.840.1.113883.3.117.1.7.1.207)

Comfort Measures
Comfort Measures are represented with the QDM datatypes and value sets of:
- Intervention, Order: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)
- Intervention, Performed: Comfort Measures (OID: 1.3.6.1.4.1.33895.1.3.0.45)

Emergency Department Visit
- Emergency Department Visit is represented with the QDM datatype and value set of Encounter, Performed: Emergency Department Visit (OID: 2.16.840.1.113883.3.117.1.7.1.292)

To access the value sets for the measure, please visit the Value Set Authority Center (VSAC), sponsored by the National Library of Medicine, at this link: https://vsac.nlm.nih.gov/.

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable, the measure is not stratified.

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
Not applicable

S.16. Type of score:
Rate/proportion
If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. Calculation Algorithm/Measure Logic *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
See attached HQMF file

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
eMeasures are calculated using only the structured data collected in certified EHR technology (CEHRT). Data not present in the structured field from which the measure draws will not be included in the measure calculation.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Hospitals report EHR data using Certified Electronic Health Record Technology (CEHRT), and by submitting Quality Reporting Document Architecture Category 1 (QRDA-1).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF0441_CMS102v3_STK10_Bonnie_Testing.xlsx,STK10_eCQM_testing_attachment.docx

Measure Number (*if previously endorsed*)**:** 2837

Measure Title: STK-10: Assessed for Rehabilitation

Date of Submission: 1/14/2016

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	X Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the sub criteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure*. *Testing must be provided for all the sources of data specified and intended for*

measure implementation. ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒other: Bonnie Test Cases	☒ other: Bonnie Test Cases (see question 1.2)

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

This measure is a legacy eCQM (an NQF endorsed measure that has been re-specified into an eMeasure and is currently used in a federal quality program/programs). In accordance with the modified NQF requirements for testing of eMeasures, the Bonnie testing tool was used to simulate a testing environment where measure specifications and HQMF output are tested against synthetic test data. Measure developers rely on the results in Bonnie to confirm whether the measure logic is performing as expected. Reference the eCQI Resource Center website (https://ecqi.healthit.gov/ecqm-tools/tool-library/bonnie) or the Bonnie testing tool website (https://bonnie.healthit.gov/) for more information about Bonnie functionality and its role in measure development. Please also reference the Bonnie testing worksheet attachment for detailed Bonnie test cases and testing results for this measure.

While Bonnie does consume HQMF specifications, we are not able to test the measure with data from HQMF implemented in EHRs.

1.3. What are the dates of the data used in testing? The measure specifications were tested in the Bonnie testing environment which mimics the year 2012.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☒ other: Bonnie Test Cases

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Not applicable.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

24 unique synthetic patient records were created in the BONNIE testing system for this measure. Cases were used to test the validity of each data element and timing relationship in the measure. Patient characteristics such as age, diagnosis, and discharge status were pre-determined to provide a variety of scenarios that adequately tested for patients passing each data element and failing each data element. Data included in cases and tested for this measure included diagnoses, discharge statuses, patient orders, rehabilitation assessment and therapy documentation, age, length of stay, and ED and inpatient encounters.

All 24 cases passed or failed as expected based on the data included in the case, confirming the measure logic is accurate and valid.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Not applicable. Bonnie test patients were developed for use across the different aspects of testing.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Not applicable.

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*May be one or both levels*)

X **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

The chart-abstracted version of this measure has been in national use since the 4th quarter of 2009. At the time the chart-abstracted measure was originally tested, extensive tests of measure reliability were conducted. At present, no performance data for the electronic version of the measure are currently available. Below is the reliability testing summary for NQF #0441 STK-10 Assessed for Rehabilitation, from which this measure is derived.

Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with monthly data transmission to The Joint Commission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of approximately 900 patient records identified at 30 pilot sites over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rates of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify potential measure enhancements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly

basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures.) The following is a list of the major tests conducted on the submitted data for this measure:

- Transmission of complete data
- Usage of data received (to understand if the healthcare organization provides the relevant service to treat the relevant population)
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity
- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors is as follows:

- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data element agreement rates reported to The Joint Commission for the chart-abstracted version of this measure for the time period of one (4Q2010 – 3Q2011) year have shown an overall agreement rate of 98.3% . This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-10.

Data Elements	Total 'n' numerator	Total 'n' denominator	Rate
Assessed for Rehabilitation			
Services	513	528	97.2%
Clinical Trial	712	714	99.7%
Comfort Measures Only	709	714	99.3%
Elective Carotid Intervention	711	714	99.6%

These agreement rates are considered to be well within acceptable levels.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Agreement rates for individual data elements tested for the chart-abstracted version of this measure were within acceptable levels. Once data are available for analysis, it is expected that reliability tests of the eCQM version of this measure will yield similar results.

.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The Bonnie testing tool and environment were used to test the measure logic and value sets. Each data element and logic statement was tested to confirm actual results met expectations. Bonnie testing includes negative and positive testing of each data element in the measure. Positive testing ensures patients expected to be included in the measure are included. Negative testing ensures that patients who do not meet the data criteria are not included in the measure. An example of negative testing would be to include test cases with pediatric ages to ensure that pediatric patients are not included in the measure.

Denominator test cases positively test to ensure patients with principal diagnosis of ischemic or hemorrhagic stroke pass and fall into the measure. We have created negative test cases, testing to ensure patients without a principal diagnosis of ischemic or hemorrhagic stroke do not fall into the denominator.

Numerator test cases positively test to ensure patients who were either assessed for rehabilitation, received rehabilitation therapy, or were discharged to rehab are included in the numerator. Negative test cases ensure that a patient who did not receive rehabilitation services do not pass the numerator.

Denominator exclusion and exception test cases for this measure ensure that patients are properly removed from the denominator if they have comfort measures, specific discharge statuses other than discharged home. Negative test cases are also run. For example, cases with comfort measures ordered but not within the specified time frame are expected to fail, and should not be removed from the denominator. Testing confirmed patients meeting the exclusion and exception criteria are removed from the measure appropriately, while those that do not meet the criteria are retained in the denominator population.

A review of the measure specifications was also conducted to confirm the logic was properly expressed within the current version of the QDM and confirmed the logic matches the clinical intent of the measure, as stated in the measure header. This is done through use of a logic checklist to facilitate review of the measure logic, according to several checks, a few examples of which are included below:

- Is the intent of the measure described in the measure description articulated/ captured in the measure logic?
- Do the logic elements map to definitions in the measure narrative, data dictionary or supporting reference documentation?
- Do the populations in the narrative align with the populations defined in the logic?
- Are the mathematic inequalities reflective of the measure intent and represent the intended populations (for example: when intended the inequality represents less than rather than less and or equal to)?

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Bonnie results provide coverage and passing rates. This measure reached 100% coverage, confirming there is a test case for each pathway of logic. This measure also has a 100% passing rate, confirming all test cases performed as expected.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

This measure performed as expected in Bonnie.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Exclusions in the eCQM align with the chart-based version of the measure, and are clinically necessary for the interpretation of the measure. As noted previously, the chart-abstracted version of this measure has been in national use since the 4th quarter of 2009, and no data are available for the eCQM version of the measure. The below analysis addresses exclusions testing performed for the chart-abstracted version of the measure from which this measure is derived.

The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patient who have a Length of Stay greater than 120 days
- Patients with Comfort Measures
- Patient enrolled in a Clinical Trial
- Patients admitted for Elective Carotid Intervention
- Patients discharged to another hospital
- Patients who left against medical advice
- Patients who expired
- Patients discharged to home for hospice care
- Patients discharged to a health care facility for hospice care

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are given below.

Exclusion: Comfort Measures Only
Overall Occurrence n = 163,225
Overall Occurrence Percentage: 10.4%
Minimum: 0.30%
10th Percentile: 5.26%
Median: 10%
90th Percentile: 15.8%
Maximum: 35.7%

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%
10th Percentile: 0.24%

Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for Elective Carotid Intervention
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.5%
90th Percentile: 25.3%
Maximum: 95.2%

Exclusion: Discharge Disposition - Patients discharged to another hospital
Overall Occurrence n = 449,924
Overall Occurrence Percentage: 35.7%
Minimum: 0.787%
10th Percentile: 25%
Median: 35.4%
90th Percentile: 46%
Maximum: 76.2%

Exclusion: Discharge Disposition - Patients who left against medical advice
Overall Occurrence n = 8,396
Overall Occurrence Percentage: 0.67%
Minimum: 0.067%
10th Percentile: 0.34%
Median: 0.86%
90th Percentile: 2.4%
Maximum: 9.67%

Exclusion: Discharge Disposition - Patients who expired
Overall Occurrence n = 76,168
Overall Occurrence Percentage: 6.04%
Minimum: 0.398%
10th Percentile: 1.9%
Median: 5.08%
90th Percentile: 10.2%
Maximum: 20.1%

Exclusion: Discharge Disposition - Patients discharged to home for hospice care
Overall Occurrence n = 658,264
Overall Occurrence Percentage: 52.2%
Minimum: 6.25%
10th Percentile: 39%
Median: 51.9%
90th Percentile: 64%
Maximum: 94.3%

Exclusion: Discharge Disposition - Patients discharged to a health care facility for hospice care
Overall Occurrence n = 37,804
Overall Occurrence Percentage: 3%

Minimum: 0.169%
10th Percentile: 0.86%
Median: 3.01%
90th Percentile: 6.4%
Maximum: 23%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Analysis of these data for the chart-abstracted progenitor of this measure indicated that all exclusions were appropriate. It is believed that results for exclusions in the eCQM will be similar when sufficient data have been received to perform such an analysis.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

- ☒ **No risk adjustment or stratification**
- ☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
- ☐ **Stratification by** Click here to enter number of categories **risk categories**
- ☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

A similar methodology will be used for the eCQMs.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

NQF#0441: STK-10 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10th Percentile
3Q2011	154	0.95859	0.1098	1	1	1	0.95833	0.88889
2Q2011	155	0.95968	0.09234	1	1	1	0.95918	0.9
1Q2011	160	0.95387	0.09443	1	1	1	0.94595	0.85165
4Q2010	136	0.95043	0.11751	1	1	1	0.95339	0.85714
3Q2010	131	0.95645	0.07732	1	1	0.99038	0.95	0.87059
2Q2010	122	0.95388	0.0842	1	1	0.99403	0.94444	0.88119
1Q2010	99	0.94899	0.0808	1	1	0.98333	0.93421	0.84615
4Q2009	49	0.95212	0.07659	1	1	0.97744	0.95	0.85294

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

It should be noted that since data collection on this measure is completely voluntary for both The Joint Commission and PCNASR, the hospitals reporting on this measure are self-selected, and therefore, presumably have a particular interest and concern for improving stroke care. For this reason, the measure results demonstrated by this group most likely significantly overstates the rate for the population of all health care organizations.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
***If only one set of specifications, this section can be skipped*.**

Note*: This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Multiple data sources are not used for this measure.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i*.e., what do the results mean and what are the norms for the test conducted*)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Data not present in the structured field from which the measure draws will not be included in the measure calculation. In the Bonnie testing environment, missing data are tested as an expected "Fail" for that data element, and actual

performance (whether or not the case fails) is compared to expectations to ensure missing data impacts data element and overall measure calculation as expected. This is the extent of missing data analysis that can be performed in Bonnie.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

See response for question 2b7.1 above.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

The measure performed as expected in the Bonnie testing environment.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment **Attachment:** STK10_eCQM_NQF_Measure_Feasibility_Assessment_Report.docx

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data**

collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Not applicable

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
Value sets are housed in the Value Set Authority Center (VSAC), which is provided by the National Library of Medicine (NLM), in coordination with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

Viewing or downloading value sets requires a free Unified Medical Language System® (UMLS) Metathesaurus License, due to usage restrictions on some of the codes included in the value sets. Individuals interested in accessing value set content can request a UMLS license at (https://uts.nlm.nih.gov/license.html)

There are no other fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Public Health/Disease Surveillance	Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs EHR Incentive Program https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms Hospital Accreditation Program http://www.jointcommission.org

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.

Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 hospitals (2015)

Name of program and sponsor: EHR Incentive Program; Centers for Medicare & Medicaid Services
Purpose: The American Recovery and Reinvestment Act of 2009 (ARRA) (Pub.L. 111–5) was enacted on February 17, 2009. Title IV of Division B of ARRA amends Titles XVIII and XIX of the Social Security Act (the Act) by establishing incentive payments to eligible professionals (EPs), eligible hospitals, and critical access hospitals (CAHs), and Medicare Advantage Organizations to promote the adoption and meaningful use of interoperable health information technology (HIT) and qualified electronic health records (EHRs). These incentive payments are part of a broader effort under the HITECH Act to accelerate the adoption of HIT and utilization of qualified EHRs.
Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4,847 active registrations (September, 2015)

Name of program and sponsor Hospital Accreditation Program; The Joint Commission
Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
Geographic area and number and percentage of accountable entities and patients included Nationwide; 3300 Joint Commission-accredited hospitals (2014)

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

While the chart-abstracted version of this measure has indicated improvement over time, this measure is a legacy eCQM that is currently in use in the Hospital Inpatient Quality Reporting Program (HIQR) and the EHR Incentive Program. At present, no performance data for the electronic version of the measure are currently available.

In CY2016, CMS is requiring organizations participating in HIQR to electronically submit 1 quarter of data on 4 of 28 available eCQMs. This measure is one of the 28. Data will be accepted through February 28th, 2017.

For more information, refer to CY2016 IPPS Final Rule, located here: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY2016-IPPS-Final-Rule-Home-Page-Items/FY2016-IPPS-Final-Rule-Regulations.html

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Not applicable

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0244 : Stroke and Stroke Rehabilitation: Rehabilitation Services Ordered
0441 : STK-10: Assessed for Rehabilitation

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0244 : Stroke and Stroke Rehabilitation: Rehabilitation Services Ordered- no longer NQF endorsed

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
NQF#0244 focuses on rehabilitation orders written prior to hospital discharge and not the rehabilitation assessment or services received by the patient. NQF#0441:STK-10 Assessed for Rehabilitation. The measures are completely harmonized to the extent possible, given the fact that the data source for #0441 is the paper medical record, and the data source for #2837 is the electronic health record.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Although both NQF#2837 STK10: Assessed for Rehabilitation and NQF#0244 target ischemic or hemorrhagic stroke patients in the acute inpatient setting, NQF#2837 is superior for two reasons. First, the numerator statement for NQF#2837 is a broader measure of quality and encompasses the total ischemic or hemorrhagic stroke inpatient population. The proportion of ischemic or hemorrhagic stroke patients who are assessed for or receive rehabilitation services during the acute inpatient hospitalization are captured in the numerator population. Patients must be assessed before services can be ordered. Rehabilitation services may be ordered but not implemented. Consequently, rehabilitation services are not received when orders are not carried out. NQF#2837 includes stroke patients who receive rehabilitation services in the numerator population. Second, NQF#0244 focuses on

rehabilitation orders written prior to hospital discharge, but capture these data after hospital discharge in the outpatient setting.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0434
Measure Title: STK-01: Venous Thromboembolism (VTE) Prophylaxis
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of ischemic or hemorrhagic stroke patients who received VTE prophylaxis or have documentation why no VTE prophylaxis was given on the day of or the day after hospital admission.
This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-2: Discharged on Antithrombotic Therapy, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-4: Thrombolytic Therapy, STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-6 Discharged on Statin Medication, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs.
Developer Rationale: Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Prevention of venous thromboembolism (VTE) through the use of prophylactic treatment for high-risk stroke patients is recommended by both the American Heart Association/American Stroke Association (AHA/ASA) and the American College of Chest Physicians (ACCP). Data suggest that VTE prophylaxis should be administered to stroke patients with restricted mobility soon after hospitalization. Results of studies evaluating the efficacy of VTE prophylaxis provide consistent findings of reductions in event rates by 50%-75% in most studies. These findings are quite consistent across hundreds of clinical trials and for many patient populations; placebo control groups are no longer considered ethical (Leizorovicz A, 2004).
Healthcare organizations that track VTE prophylaxis for internal quality improvement purposes have seen significant improvement in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

Numerator Statement: Ischemic or hemorrhagic stroke patients who received VTE prophylaxis or have documentation why no VTE prophylaxis was given on the day of or the day after hospital admission.
Denominator Statement: Ischemic or hemorrhagic stroke patients
Denominator Exclusions:

- Less than 18 years of age
- Length of Stay < 2 days
- Length of Stay > 120 days
- Comfort measures only documented on day of or day after hospital arrival
- Enrolled in clinical trials related to stroke
- Admitted for elective carotid intervention

Measure Type: Process
Data Source: Electronic Clinical Data, Paper Medical Records
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2008 **Most Recent Endorsement Date:** Nov 01, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Summary of prior review in 2012:

Results of studies evaluating the efficacy of VTE prophylaxis following stroke provide consistent findings of reductions in event rates by more than 70% (Geerts, 2001). These findings are consistent across multiple clinical trials. 2007 AHA/ASA Guidelines for the Early Management of Adults with Ischemic Stroke:

- Class I, Level of Evidence A - *Subcutaneous administration of anticoagulants is recommended for treatment of immobilized patients to prevent deep vein thrombosis. The ideal timing for starting these medications is not known.*
- Class IIa, Level of Evidence B - *The use of intermittent external compression devices is recommended for treatment of patients who cannot receive anticoagulants*.

Antithrombotic Therapy and Prevention of Thrombosis, 9th Ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines, page 33S-34S: Grade 2B -*In patients with acute ischemic stroke and restricted mobility, we suggest prophylactic dose SC UFH or LMWH or intermittent pneumatic compression devices over no prophylaxis.*

The 2012 Committee questioned why—given that graduated compression stockings have not been shown to reduce VTE risk or death in the first seven days post-stroke—the measure isn't limited to chemoprophylaxis. The developer clarified that although graduated compression stockings (i.e., TED hose) are not sufficient for VTE prophylaxis, other mechanical devices such as pneumatic or sequential compression devices are considered appropriate for patients who are not eligible for chemoprophylaxis (although the developer acknowledged that the evidence for mechanical devices is not as strong).

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☐ **The developer provided updated evidence for this measure:**

Guidance from the Evidence Algorithm

Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency high(Box 5) → Box 5a High

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and*

there is no need for repeat discussion and voting on Evidence?

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following national trend data:

	CY 2010	CY2011	CY 2012	CY 2013	CY 2014
# hospitals	138	157	158	264	1299
# cases (den)	25,031	29,680	29,538	44,644	213,000
National aggregate rate	0.88	0.92	0.94	0.96	0.97
Mean hospital rate	0.83	0.88	0.92	0.95	0.96
50th percentile	0.88	0.94	0.96	0.98	0.99
10th and 90th percentiles	0.60 (10th) 0.98 (90th)	0.67 (10th) 0.996 (90th)	0.82 (10th) 0.99 (90th)	0.88 (10th) 0.99 (90th)	0.91 (10th) 1.0 (90th)

Participation in this measure has grown significantly in the past five year. Hospital performance national seems to be very high with little room for improvement.

In 2012 the Committee suggest that performance rates be reported by pharmacological vs. mechanical treatment, and for ischemic vs. hemorrhagic stroke patients.

Disparities

The developer does not provide disparities data from use of this measure. Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age."

Questions for the Committee:

- *Data for 2014 includes 1299 hospitals. What proportion of hospitals caring for stroke patients is captured in this data?*
- *Data over five years shows significant improvement. How much further improvement in performance is likely using this measure?*
- *Is there a gap in care that warrants a national performance measure?*
- *How can this measure be used to better understand disparities in care and outcomes for certain groups?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **There is strong and compelling evidence to support the the linkage between VTE prophylaxis and improved outcomes in patients with limited mobility after stroke. The extension to all patients, as presented in the measure, has much weaker evidence. For example, the PREVAIL study limited recruitment to individuals with limited mobility, but the results were extrapolated to stroke generally.

**not aware of any new studies/information

1b. Performance Gap

Comments: **A performance gap was identified in the literature. As a maintenance measure, this measure has real world results: Hospitals that adopted the measure demonstrated a performance gap initially and improvement over time.

**performance data provided, high participation in this measure however there is no disparities data from use of the measure. If hospitals could provide disparities data, this would help further improvement.

1c. High Priority (previously referred to as High Impact)

Comments: **Not applicable.

**NA

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): medical record abstraction (paper or electronic). This is not an eMeasure.

Specifications:

- Four data elements define the numerator and seven data elements define the denominator.
- Developer report that "Since the last endorsement date, the data element Reason for Oral Factor Xa was added to the measure logic and the algorithm revised based on Nov. 4, 2011 FDA approval of rivaroxaban for stroke prevention in patients with atrial fibrillation. Direct oral anticoagulant agents have been added to Appendix H, Table 2.1 VTE Prophylaxis Inclusion Table as they have been approved for the U.S. Food and Drug Administration."
- Diagnosis and procedure codes have been converted from ICD-9 to ICD-10-CM diagnosis and ICD-10-PCS procedure codes. See attached spreadsheet Appendix A.1 Table 8.1 (ischemic stroke) and 8.2 (hemorrhagic stroke).
- A Conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications:
    - "The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015 developer advisory panel determined that the intent of the measure was changed by the conversion to ICD-10 and a crosswalk was posted."
- Sampling monthly or quarterly is allowed; instructions for sampling are provided.
- A web-based data collection tool has been developed but is not described. Hospitals are not required to use this tool.
- The measures is specified at the hospital level of analysis.

In the 2012 evaluation a Committee member questioned why the measure allows for an exclusion due to patient refusal. Other members noted that patients may refuse mechanical devices because they are uncomfortable.

Questions for the Committee :

- *Are the changes to the measure specifications appropriate?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:
Inter rater-reliability of the data elements for one year (4Q2010-3Q2011) in 77 hospitals and 739 patient records showed an overall agreement rate of 97.84%. Two data elements found less than 90% agreement: Reason for no VTE prophylaxis (82.2%) VTE prophylaxis (89.9%). No Kappa scores were presented.

Describe any updates to testing – no new information provided

SUMMARY OF TESTING
Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**
Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing see above

Results of reliability testing see above

Guidance from the Reliability Algorithm
Precise specifications (Box 1) →empirical testing as specified (Box 2) → reliability testing with patient-level data elements (Box 8) → assessed all critical data elements (Box 9)→ high/moderate certainty the data elements are reliable (Box 10) → Moderate (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Questions for the Committee:

- *The developer has not provided any new reliability testing. Does the Committee agree there is no need for repeat discussion and voting on Reliability?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.
Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

In the 2012 evaluation the Committee questioned why both mechanical and pharmacological responses are included in the measure given the difference in evidence. The developer explained that they wanted to construct a

measure that would assess provision of appropriate therapy in all patients, and noted that pharmacological treatments are contraindicated for hemorrhagic stroke patients in the first several days post-stroke.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

1. Face validity of the data was assessed via survey and focus groups of hospitals participating in the pilot test. The developer reports that "All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant." The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.
2. Ongoing feedback from measure users that is systematically reviewed to identify trends and/or measure specifications that require clarification or revision. This measure receives more questions than other measures in the stroke set, resulting from modification to the measure following CMS alignment with similar Surgical Care Improvement Project (SCIP) VTE and VTE core measures.

Describe any updates to validity testing – New empirical validity testing data is presented

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

Pearson Correlation Coefficients were calculated for seven stroke measures for hospitals that reported 12 months of data and had 30 or more denominator cases for the year (2014-2015).Presumably high performing hospitals would do well on all stroke performance measures.

Validity testing results:

Analysis of 1,318 hospitals and 2, 206,379 patients records generated a table of Pearson Correlation Coefficient results that shows a statistically significant (P<000.1), positive correlation of this measure with six other stroke performance measures supporting the hypothesis that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

Patients are excluded from the measure if they are under age 18, have a length of stay greater than 120 days, are receiving comfort measures only as of the day of or day after admission, enrolled in clinical trials, or are discharged within 24 hours.

New data on the frequency of exclusions is presented for 2, 206,379 admissions in 1,318 hospitals:

- Overall exclusion for Comfort Measures Only = 4.46% (range 0.31 – 30.8%)
- Overall exclusions for elective carotid intervention = 11.8% (range 0.16-95.3%)

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:
Performance data is presented above under opportunity for improvement. Complete details of the data is presented in 1b.

The developer explains their approach: "There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations."

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods: not applicable

2b7. Missing Data

The developer describes the handling of missing data in the measure specifications (S.22) but does not provide information on the frequency of missing data or potential impact on measure results.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → potential threats to validity mostly assessed (Box2) →empirical validity testing (Box 3) → measure score testing (Box 6) → sound conceptualization and hypotheses (Box 7) →high confidence that score are a valid indicator of quality(Box 8a) → High

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications
Comments: **The specification measures all patients with ischemic or hemorrhagic stroke, however the evidence reviewed specified stroke patients with low mobility.
**specifications consistent with the evidence
2a2. Reliability Testing
Comments: **The measure addresses the evidence effectively however it extends the treatment beyond the evidence by not addressing the issue of mobility. Several RCTs and reviews support leave out the mobile stroke patient from studies of VTE prophylaxis.
**new validity testing demonstrates sufficient validity
2b2. Validity Testing
Comments: **Documentation of why no VTE prophylaxis is considered a positive response and is a quite permissive criterion.
**No

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments: **The measure appears to be reliable, with good results for inter-rater reliability with a large N (77 hospitals, 739 patients). Overall agreement rate = 98%. Reliability was tested at the data element level.
**No need for new reliability testing

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is medical record abstraction: not all hospitals are able to generate the data electronically so a paper-based option is available
- Some data elements are in electronic form
- The measure is in use in the Joint Commission's Stroke Certification program.
- Data collection burden is significant.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments: **The measure is quite feasible as evidenced by prior adoption.
**Multiple ways that data collection is implemented- no concerns.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**
OR
Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details

- Public reporting: The measure is publicly reported on The Joint Commission's Quality Check http://www.qualitycheck.org/consumer/searchQCR.aspx and CMS's Hospital Compare https://www.medicare.gov/hospitalcompare/search.html
- Payment program: The measure is part of the Hospital Inpatient Quality Reporting Program (Hospital IQR)

Improvement results See data under gap above. The developer summarizes "The rate of VTE prophylaxis has steadily increased over the past five years based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 97.3%. A gap of 9-10% still exists for the 10th percentile of hospitals. Modest differences in VTE prophylaxis in black patients have been reported by GWTG (Schwamm, 2010)."

Unexpected findings (positive or negative) during implementation None reported

Potential harms None reported.

Feedback loop:

- Ongoing feedback from end users to the measure developer.

***Questions for the Committee**:*

- o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences
Comments: **The measure is usable as evidenced by prior adoption, in CY2014 it was adopted at 1,299 hospitals for 213,000 cases. This is 23% of 5627 hospitals in the US.
**Accountability is high- public reporting and Hospital IQR

Criterion 5: Related and Competing Measures

Related or competing measures

- **0240 Stroke and Stroke Rehabilitation: Deep Vein Thrombosis (DVT) Prophylaxis for Ischemic Stroke or Intracranial Hemorrhage (AMA-PCPI) –** competing measure at clinician level of analysis
- **0239 Venous Thromboembolism (VTE) Prophylaxis (AMA-PCPI)** – for surgical patients
- **0371 Venous Thromboembolism Prophylaxis (TJC)** – identical to 0434 but excludes stroke patients

Harmonization

In the prior evaluation the Committee expressed a desire for a single measure that could be used at the clinician and facility levels. The two developers agreed it was not feasible at that time to create a single measure. The Committee recommended continued and aggressive efforts for harmonization when possible, and requested an update on progress on harmonization at the time of annual review.

The developer reports that:

- Measures NQF# 0371 specifically excludes the stroke population. The measures are completely harmonized in terms of measure specifications and data element definitions.
- Measure 0239 is a physician performance measure with a targeted population of surgical patients identified through CPT codes and thus is a different level of measurement. This measure evaluates physician practice as opposed to hospital processes.

Questions for the Committee:

- *Are measures 0434 and 0240 optimally harmonized to reduce measurement burden for providers?*

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NQF #: 0434 NQF Project: Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The focus of the measure is to prevent venous thromboembolism (VTE) following ischemic or hemorrhagic stroke. VTE prophylaxis treatment >> VTE prevention >>decreased morbidity and mortality.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
The central topic for the measure is prevention of venous thromboembolic events by use of prophylaxis in hospitalized patients with ischemic or hemorrhagic stroke. Patients with restricted mobility, as is the case for the majority of hospitalized stroke patients, are particularly high-risk for developing VTE. This measure is consistent with the body of evidence that supports guideline recommendations from the American Heart Association/American Stroke Association for secondary stroke prevention, and the American College of Chest Physicians (ACCP)for antithrombotic therapy and prevention of thrombosis.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: There are literally hundreds of well designed, prospective, randomized clinical trials published over the past 35 years that have clearly demonstrated the efficacy of both pharmacologic and mechanical forms of VTE prophylaxis to prevent VTE events (including asymptomatic, symptomatic, and deaths from VTE). In 2008, the American College of Chest Physicians (ACCP) summarized more than 700 articles, most of which represented published clinical trials on prevention of VTE events. Similarly, there are dozens of observation studies that have evaluated use of VTE prophylaxis in hospitalized patients (usually through medical record audit) which show consistent underuse of prophylaxis.

Several systematic reviews of the evidence as it relates to the prevention of venous thromboembolism in stroke patients have been conducted. The Cochrane Stroke Group recently updated a review of low-molecular-weight heparins or heparinoids versus standard unfractionated heparin for acute ischemic stroke (2011), originally published in 1995 with a second update in 2005. The search strategy included the Cochrane Stroke Group's Trials Register (last searched June 2007); the Cochrane Central Register of Controlled Trials (CENTRAL)(The Cochrane Library 2007, Issue 2): MEDLINE (1966 to June 2007); and EMBASE (1980 to June 2007). For the previous version of this review, MedStrategy (1995) was also searched and pharmaceutical companies contacted. Selection criteria were specific for randomized trials comparing heparinoids or low-molecular-weight heparins with standard unfractionated heparin in persons with acute ischemic stroke.

This review identified nine trials involving 3137 ischemic stroke patients. Four trials compared a heparinoid (danaparoid) with standard unfractionated heparin. Another four trials compared enoxaparin or certoparin with unfractionated heparin, and one compared an unspecified low-molecular-weight heparin. Low-molecular-weight heparins or heparinoids were associated with a significant reduction in the odds of VTE compared with standard unfractionated heparin (odds ratio (OR) 0.55; 95% CI 0.44 to 0.70). Three new studies

(n=2397 participants) included in this review provided no information to change previous conclusions about the effectiveness of low-molecular-weight heparins for the prevention of VTE in the ischemic stroke population.

A second recent review from the Cochrane Stroke Group selected randomized control trials comparing mechanical forms of prophylaxis for the prevention of VTE post-stroke (2010). For this review, the Cochrane Stroke Group's Trials Register (last searched November 2009); the Cochrane Central Register of Controlled Trials (CENTRAL)(The Cochrane Library 2009, Issue 4): MEDLINE (1966 to November 2009); and EMBASE (1980 to November 2009), CINAHL (1980 to November 2009), and The British Nursing Index (1985 to November 2009) were searched, as well as, ongoing trials registers (November 2009). Additionally, identified experts in the field were contacted as needed. Two trial of graduated compression stockings (n=2615 participants) and two smaller studies involving intermittent compression devices (n=177 participants) were included in the review. Graduated compression stockings were not found to reduce VTE risk (OR 0.88; 95% CI 0.72 to 1.08) or death (OR 1.13; 95% CI 0.87 to 1.47) within the first seven days post-stroke; however, intermittent compression devices were associated with a non-significant trend toward decreased risk of VTE (OR 0.45; 95% CI 0.19 to 1.10) with no impact on deaths (OR 1.04; 95% CI 0.37 to 2.89).

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting thromboprophylaxis to prevent VTE complications and reduce mortality after stroke is high. Data overwhelmingly support that thromboprophylaxis reduces VTE events in the stroke population, and that pharmacological prophylaxis is more beneficial than mechanical forms. Studies have also shown that fatal PE is prevented by thromboprophylaxis. With respect to complications of thromboprophylaxis, randomized clinical trials have demonstrated little or no increase in the rates of symptomatic intracranial and major extracranial hemorrhage (1-2%) with prophylactic doses of pharmacologic VTE prophylaxis. Risk of gastrointestinal hemorrhage or other major hemorrhage is the primary concern with ongoing therapy, although the studies have demonstrated that the risk is low and absolute benefits significantly outweigh the risks.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* Results of studies evaluating the efficacy of VTE prophylaxis following stroke provide consistent findings of reductions in event rates by more than 70% (Geerts, 2001). These findings are quite consistent across multiple clinical trials.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
There is strong evidence on the immediate hazards and net benefits of prophylactic interventions for the prevention of venous (VTE) thromboembolism. Stroke patients have a high risk of deep vein thrombosis (DVT) in the paretic or paralyzed lower extremity with a pooled DVT incidence of 55%. Approximately 5-10% of early deaths following stroke are attributed to pulmonary embolism (PE). Among 421 patients admitted to a stroke rehabilitation unit, routine duplex ultrasonography detected proximal DVT in 14% at entry to the unit, and an additional 5% of patients without DVT prophylaxis were subsequently found to have proximal DVT during the rehabilitation stay (Desmukh M, 1991).

Routine use of thromboprophylaxis reduces adverse patient outcomes while at the same time decreasing overall costs. The PREVAIL (Prevention of VTE after acute ischemic stroke with LMWH and UFH) evaluated the economic impact of enoxaparin versus unfractionated heparin for venous thromboembolism prophylaxis in patients with acute ischemic stroke. Using a decision-analytic model to analyze hospital-based costs, Pineo and associates (2012) found that total hospital cost was lower with enoxaparin versus UFH ($782 versus $922, respectively; savings $140). The higher drug cost of enoxaparin was offset by the reduction in VTE events as compared to the use of UFH. Hospital cost-savings were greatest in patients with more severe strokes (i.e., National Institutes for Health Stroke Scale (NIHSS) scores > 14).

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: American College of Chest Physicians. During our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, and provided information that was accessible and met the requirements set out in this measure maintenance form.

1c.11 System Used for Grading the Body of Evidence: GRADE

1c.12 If other, identify and describe the grading scale with definitions:

1c.13 Grade Assigned to the Body of Evidence: Grade 2B

1c.14 Summary of Controversy/Contradictory Evidence: There is no dispute that graduated compression stockings are not sufficient thromboprophylaxis for acutely ill stroke patients. Abundant data support the use of pharmacological thromboprophylaxis for acute ischemic stroke over intermittent compression devices, unless contraindicated. Data also suggest that low-molecular-weight heparin enoxaparin has a superior clinical profile in comparison with unfractionated heparin. Yet, the relative advantages and disadvantages of various prophylactic drugs remains questionable.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- Caprini JA, Arcelus JI. State-of the art venous thromboembolism prophylaxis. SCOPE on Phlebology & Lymphology 1:2005, 228-240.
- Coull BM, Williams LS, Goldstein LB, et al. Anticoagulants and Antiplatelet Agents in Acute Ischemic Stroke. Report of the Joint Stroke Guideline Development Committee of the American Academy of Neurology and the American Stroke Association (a Division of the American Heart Association) Stroke. 2002;33:1934 -1942.
- Desmukh M, Bisignani M, Landau P, Orchard TJ. Deep vein thrombosis in rehabilitating stroke patients; incidence, risk factors and prophylaxis. Am J Phys Med Rehabil. 1991;30:313-316.
- Duncan et al, Stroke Rehabilitation Clinical Practice Guidelines (Stroke. 2005;36:e100-e143.)
- Geerts WH, Heit JA, Clagett GP, Pineo GF, Colwell CW, Anderson FA, Wheeler HB. Prevention of venous thromboembolism. Chest. 2001;119:132S-175S.
- Kase CS, Albers GW, Bladin C, Fieschi C, Gabbai AA, O'Riordan W, Pineo GF; PREVAIL Investigators. Neurological outcomes in patients with ischemic stroke receiving enoxaparin or heparin for venous thromboembolism prophylaxis: subanalysis of the Prevention of VTE after Acute Ischemic Stroke wih LMWH (PREVAIL) study. Stroke. 2009;40(11):3532-40.
- Kase CS, Pineo GF. Prevention of venous thromboembolism after ischemic stroke. Curr Opin Pulm Med. 2008;14(5):389-96.
- Kelly J, Rudd A, Lewis RR, Coshall C, Moody A, Hunt BJ. Venous thromboembolism after acute ischemic stroke: a prospective study using magnetic resonance direct thrombus imaging. Stroke. 2004;35:2320-2325.
- Kelly J, Rudd A, Lewis R, Hunt BJ. Venous thromboembolism after acute stroke. Stroke. 2001;32:262-267.
- Leizorovicz A, Mismetti P. Preventing venous thromboembolism in medical patients. Circulation. 2004 Dec 14;110(24):13-19.
- Michota FA. Venous thromboembolism prophylaxis in medical patients. Curr Opin Cardiol. 2004 Nov;19(6):570-4.
- Naccarato M, Grandi FC, Dennis M, Sandercock PAG, Cochrane Stroke Group. Physical methods for preventing deep veni thrombosis in stroke. The Cochrane Library. 2010 Aug 4;(3): CD001922.
- Pineo G, Lin J, Stern L, Subrahmanian T, Annemans L. Economic impact of enoxaparin versus unfractionated heparin for venous thromboembolism prophylaxis in patients with acute ischemic stroke: a hospital perspective of the PREVAIL trial. J Hosp Med. 2012;7(3):176-82.
- Post-Stroke Rehabilitation Guideline No.16, Agency for Healthcare Policy and Research (Now known as Agency for Healthcare Research and Quality), 1995.
- Sandercock PAG, Counsell C, Tseng MC, Cochrane Stroke Group. Low-molecular-weight heparins or heparinoids versus standard unfractionated heparin for acute ischemic stroke. Cochrane Database Syst Rev. 2011 March 6;(8): CD000119.
- Vergouwen MD, Roos YB, Kamphuisen PW. Venous thromboembolism prophylaxis and treatment in patients with acute stroke and traumatic brain injury. Curr Opin Pulm Med. 2008;14(2):149-55.
- Warlow C, Ogston D, douglas AS. Deep venous thrombosis of the legs after strokes, part I: incidence and predisposing factors. BMJ. 19761:1178-1181.
- Wijdicks EF, Scott JP. Pulmonary embolism associated with acute stroke. Mayo Clin Proc. 1997;72:297-300.

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)*:
2007 AHA/ASA Guidelines for the Early Management of Adults with Ischemic Stroke, page 1689.

Class I, Level of Evidence A
Subcutaneous administration of anticoagulants is recommended for treatment of immobilized patients to prevent deep vein thrombosis. The ideal timing for starting these medications is not known.

Class IIa, Level of Evidence B
The use of intermittent external compression devices is recommended for treatment of patients who cannot receive anticoagulants.

Antithrombotic Therapy and Prevention of Thrombosis, 9th Ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines, page 33S-34S.

Grade 2B
In patients with acute ischemic stroke and restricted mobility, we suggest prophylactic dose SC UFH or LMWH or intermittent pneumatic compression devices over no prophylaxis.

In patients with acute stroke and restricted mobility, we suggest against elastic compression stockings.

In patients with acute primary intracerebral hemorrhage and restricted mobility we suggest prophylactic-dose LMWH over prophylactic dose UFH.

In patients with primary intracerebral hemorrhage and restricted mobility, we suggest against elastic compression stockings.

1c.17 Clinical Practice Guideline Citation: Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A Guideline From the American Heart Association/American Stroke, Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1689-90.

Guyatt GH, Akl EA, Crowther M, Gutterman DD, Sch?nemann HJ, and for the American College of Chest Physicians Antithrombotic Therapy and Prevention of Thrombosis Panel. Executive summary: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141: 33S-34S.

1c.18 National Guideline Clearinghouse or other URL:
http://www.guidelines.gov/content.aspx?id=10911&search=early+management+of+adults+with+ischemic+stroke

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: This guideline was approved by the American Heart Association Science Advisory and Coordinating Committee on January 6, 2007. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.23 Grade Assigned to the Recommendation: AHA: Class I, Level of Evidence A; ACCP: GRADE 2B

1c.24 Rationale for Using this Guideline Over Others: The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

The American College of Chest Physicians Guidelines are comprehensive in scope and also include recommendations for VTE prevention in patients with ischemic and hemorrhagic stroke. Guideline recommendations for pharmacological and mechanical thromboprophylaxis are combined.

Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: High 1c.26 Quality: High1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0434_Evidence_MSF5.0_Data.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Prevention of venous thromboembolism (VTE) through the use of prophylactic treatment for high-risk stroke patients is recommended by both the American Heart Association/American Stroke Association (AHA/ASA) and the American College of Chest Physicians (ACCP). Data suggest that VTE prophylaxis should be administered to stroke patients with restricted mobility soon after hospitalization. Results of studies evaluating the efficacy of VTE prophylaxis provide consistent findings of reductions in event rates by 50%-75% in most studies. These findings are quite consistent across hundreds of clinical trials and for many patient populations; placebo control groups are no longer considered ethical (Leizorovicz A, 2004).

Healthcare organizations that track VTE prophylaxis for internal quality improvement purposes have seen significant improvement in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
In October, 2009, The Joint Commission added the stroke (STK) measure set as a new core measure option to meet performance measurement requirements for Joint Commission hospital accreditation purposes. Below is the specified level of analysis for STK-1 beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 2668 denominator cases; 2220 numerator cases; 49 hospitals; 0.83208 national aggregate rate; 0.76618 mean of hospital rates; 0.22024 standard deviation; 0.96296 90th percentile rate; 0.92453 75th percentile rate/upper quartile; 0.82813 50th percentile rate/median rate; 0.69565 25th percentile rate/lower quartile; and, 0.43032 10th percentile rate .

CY 2010: 25031 denominator cases; 22083numerator cases; 138 hospitals; 0.88223 national aggregate rate; 0.83435 mean of hospital rates; 0.16701 standard deviation; 0.9837 90th percentile rate; 0.95775 75th percentile rate/upper quartile; 0.87923 50th percentile rate/median rate; 0.78261 25th percentile rate/lower quartile; and, 0.5959610th percentile rate.

CY 2011: 29680 denominator cases; 27393 numerator cases; 157 hospitals; 0.92294 national aggregate rate; 0.88278 mean of hospital rates; 0.14973 standard deviation; 0.99563 90th percentile rate; 0.97521 75th percentile rate/upper quartile; 0.94152 50th percentile rate/median rate; 0.85714 25th percentile rate/lower quartile; and, 0.66667 10th percentile rate.

CY 2012: 29538 denominator cases; 27859 numerator cases; 158 hospitals; 0.94316 national aggregate rate; .092953 mean of hospital rates; 0.09568 standard deviation; 1.0 90th percentile rate; 0.98851 75th percentile rate/upper quartile; 0.95652 50th percentile rate/median rate; 0.90909 25th percentile rate/lower quartile; and, 0.81619 10th percentile rate.

CY 2013: 44644 denominator cases; 42851 numerator cases; 264 hospitals; 0.95984 national aggregate rate; 0.94689 mean of hospital rates; 0.10083 standard deviation; 1.0 90th percentile rate; 0.99442 75th percentile rate/upper quartile; 0.9754 50th percentile rate/median rate; 0.94306 25th percentile rate/lower quartile; and, 0.8764 10th percentile rate.

CY 2014: 213000 denominator cases; 207132 numerator cases; 1299 hospitals; 0.97254 national aggregate rate; 0.96243 mean of

hospital rates; 0.07648 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.98618 50th percentile rate/median rate; 0.95853 25th percentile rate/lower quartile; and, 0.90909 10th percentile rate.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
Not applicable.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable. According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al. 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistical figures reveal that although women have a higher "life-time risk of stroke" than men, they have a lower "age-adjusted risk of stroke" and, women ages 45-85 have a lower overall rate of stroke. Stroke risk significantly increases for women > 85 years old; however, women live longer than men which may account in part for the difference (Roger VL, et al., 2012).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, 2011).

In the national REGARDS cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years if age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age : RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

There are few data published on racial differences in the use of VTE prophylaxis. A single study from a university hospital in Hawaii showed that Japanese patients were less likely to receive VTE prophylaxis than other racial groups. However this is not likely representative of practice in most US hospitals and may be appropriate based on Japanese patients´ lower risk for VTE events. There are robust data on racial differences in rates of VTE based on race/ethnicity. In most studies African American men have higher rates of VTE than white patients. The lowest rates of VTE are seen in patients of Asian/Pacific Islander decent.

As previously mentioned, ethnic minorities suffer higher mortality and higher rates of more severe hemorrhagic strokes. Increased stroke severity and immobilization increase the risk of developing VTE. Although no data were found noting disparities in VTE prophylaxis administration for stroke patients, there are some data that inpatient evaluation differed between African Americans and whites.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al. 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to 75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

Ischemic strokes in black patients occur earlier in life, and age-standardized mortality due to ischemic stroke are higher in blacks when compared to the general population (Schwamm, L., et al., 2010). Blacks have less access to quality care, a higher prevalence of risk factors, and more severe deficit when the stroke occurs (Qian, F., et al., 2013).

In the national REGARDS cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. Hispanics are at a greater risk for stroke as compared with non-Hispanic whites, even living in the same community (Schwamm, L., et al., 2010). A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years of age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age : RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer,et al., 2010).

In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).

Qian, F., et al., (2013) found that non-Hispanic Blacks, Hispanic, and non-Hispanic Asian Americans patients were less likely to die than non-Hispanic whites, within 30 days of admission. Then again they were less likely to die within one year of admission and within one year of discharge. In comparison with non-Hispanic white patients, the measured race/ethnicity groups had a lower 30 day mortality rate of hospital admission. Survival after one year post discharge was strongest in the non-Hispanic Asian Americans. When the authors compared non-Hispanic white groups with non-Hispanic black and Hispanic groups, the non-Hispanic black and Hispanic groups had a higher one year all cause and stroke associated re hospitalization. Non-Hispanic Asian Americans had a lower one year all cause re hospitalization rate.

There are few data published on racial differences in the use of VTE prophylaxis. A single study from a university hospital in Hawaii showed that Japanese patients were less likely to receive VTE prophylaxis than other racial groups. However this is not likely representative of practice in most US hospitals and may be appropriate based on Japanese patients´ lower risk for VTE events. There are robust data on racial differences in rates of VTE based on race/ethnicity. In most studies African American men have higher rates of VTE than white patients. The lowest rates of VTE are seen in patients of Asian/Pacific Islander decent.

In a study done by Schwamm, L., et al (2010), examiners found that there were modest differences in DVT Prophylaxis in black patients. Although small, it is thought they are large enough to result in the overall in hospital mortality due to pulmonary emboli. They continued to assess the evidence regarding the ability to ambulate at discharge, and found black and Hispanic patients were able to ambulate greater than the white population. This may be due to age or other unmeasured factors, and not related to quality of care.

As previously mentioned, ethnic minorities suffer higher mortality and higher rates of more severe hemorrhagic strokes. Increased stroke severity and immobilization increase the risk of developing VTE. Although no data were found noting disparities in VTE prophylaxis administration for stroke patients, there are some data that inpatient evaluation differed between African Americans and whites.

• Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, Yan T, Escarce JJ. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-8.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, Griffith P, Gorelick PB, Howard G, Leira EC, Morgenstern LB, Ovbiagele B, Peterson E, Rosamond W, Trimble B, Valderrama AL, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• Gelber RP, Seto TB. "Patient ethnicity and use of venous thromboembolism prophylaxis." Int J Qual Health Care. 2006; 18:23-9.
• Gillum RF, Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 2011;42(11):3294-6.
• Heit JA, Beckman MG, Bockenstedt PL, et al. "Comparison of characteristics from White- and Black-Americans with venous thromboembolism: a cross-sectional study." Am J Hematol. 2010; 85:467-71.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-75.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Jacobs BS, Birbeck G, Mullard AJ, Hickenbottom S, Kothari R, Roberts S, Reeves MJ. Quality of hospital care in African Americans and white patients with ischemic stroke and TIA. Neurology.2006;66:809-14.
• Johnston SC, Fung LH, Gillum LA, Smith WS, Brass LM, Lichtman JH, Brown AN. Utilization of intravenous tissue-type plasminogen activator for ischemic stroke at academic medical centers.: the influence of ethnicity. Stroke. 2001; 32:1061-68.
• Karve S, Balkrishnan R, Seiber E Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle

Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, Kerber KA, Skolarus LE, Meurer WJ, Burke JF, Adelman EE, Baek J, Lisabeth LD. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol. 2013;74:778–785.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Qian, F, Fonarow G C., Smith E E, et al., Racial and ehtinic diferences in outcome in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125: e78-e82.
• Schwamm L.H. Reeves, MJ, Wenqin P, et al. Race/ethnicity, quality of care, and outcomes in ischemic stroke. Circulation. 2010;121:1492-1501.
• Stein PD, Matta F. "Epidemiology and incidence: the scope of the problem and risk factors for development of venous thromboembolism." Clin Chest Med. 2010;31:611-28.
• Tang Y, Sampson B, Pack S, et al. "Ethnic differences in out-of-hospital fatal pulmonary embolism." Circulation. 2011;123:2219-25.
• White RH, Keenan CR. "Effects of race and ethnicity on the incidence of venous thromboembolism." Thromb Res. 2009;123(4):S11-7.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality, Severity of illness

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States remains attributable to stroke. More women than men die of stroke each year. Women accounted for almost 60% of US stroke deaths (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, et al., 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 ($P<0.05$).168 (US Burden of Disease Collaborators, 2013) . Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

The 2015 report by The Agency for Healthcare Research and Quality (AHRQ) indicates that
"Pharmacologic methods to prevent venous thromboembolism are safe, effective, cost-effective, and advocated by authoritative guidelines, yet large prospective studies continue to demonstrate that these preventive methods are significantly underused." Increased stroke severity and immobilization increase the risk of developing VTE. Pulmonary embolism accounts for approximately 10% of deaths after stroke, and the complication may be detected in approximately 1% of patients who have had a stroke, underscoring the importance of prevention as the most critical action step for reducing death from PE (Wijdicks and Scott, 1997). Besides being associated with life-threatening PE, symptomatic DVT also slows recovery and rehabilitation after stroke (Jauch, E.C., et al, 2013).

1c.4. Citations for data demonstrating high priority provided in 1a.3

• Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A Guideline From the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1689-90.
• Agency for Healthcare Research and Quality. Preventing Hospital Acquired Venous Thromboembolism: A guide for effective quality improvement.–Retrieved November 4, 2015 from: http://www.ahrq.gov/professionals/quality-patient-safety/patient-safety-resources/resources/vtguide/index.html.
• Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• George M, Xin T, McGruder H, Yoon P, Rosamond W, Winquist A, Hinchey J, Wall H, Pandey D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Guyatt GH, Akl EA, Crowther M, Gutterman DD, Schu?nemann HJ, and for the American College of Chest Physicians Antithrombotic Therapy and Prevention of Thrombosis Panel. Executive summary: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141: 33S.
• Heit J. "The epidemiology of venous thromboembolism in the community." Arteriosclerosis, Thrombosis and Vascular Biology 2008;28(3):370.
• Jauch EC, Saver JL, Adams HP Jr, Bruno A, Connors JJ, Demaerschalk BM, Khatri P, et al. "Guidelines for the Early Management of Patients with Acute Ischemic Stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association." [In Eng]. Stroke (Jan 31 2013).
• Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, Fang MC, Fisher M, Furie KL, Heck DV, Johnston SC, Kasner SE, Kittner SJ, mitchell PH, Rich MW, Richardson D, Schwamm LH, Wilson JA. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2014;45:241-2160-2236.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• National Heart, Lung, and Blood Institute, National Institutes of Health; reports on a conference that occurred in 1996. Stroke Belt Initiative: project accomplishments and lessons learned.
• Raskob GE, Silverstein R, Bratzler DW, Heit JA, White RH. "Surveillance for deep-vein thrombosis and pulmonary embolism: recommendations from a national workshop." Am J Prev Med. 2010: 38(4 Suppl):S502-9.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e-78-82, e-119-127.
• Stein PD, Matta F. "Epidemiology and incidence: the scope of the problem and risk factors for development of venous thromboembolism." Clin Chest Med. 2010; 31:611-28.
• Wijdicks EF, Scott JP. Pulmonary embolism associated with acute stroke. Mayo Clin Proc. 1997;72:297-300.

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input was obtained.)*
Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA), Prevention

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Safety : Venous Thromboembolism

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Appendix_A.1.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Since the last endorsement date, the data element Reason for Oral Factor Xa was added to the measure logic and the algorithm revised based on Nov. 4, 2011 FDA approval of rivaroxaban for stroke prevention in patients with atrial fibrillation. Direct oral anticoagulant agents have been added to Appendix H, Table 2.1 VTE Prophylaxis Inclusion Table as they have been approved for the U.S. Food and Drug Administration (FDA).

All ICD-9-CM diagnosis codes and ICD-9-CM procedure codes were converted to ICD-10-CM diagnosis and ICD-10-PCS procedure codes throughout the measure specifications. The Department of Health and Human Services (HHS) mandated that all entities covered by the Health Insurance Portability and Accountability Act (HIPAA) must all transition to a new set of codes for electronic health care transactions on October 1, 2015.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Ischemic or hemorrhagic stroke patients who received VTE prophylaxis or have documentation why no VTE prophylaxis was given on the day of or the day after hospital admission.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care.

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.

Four data elements are used to calculate the numerator:

• Reason for No VTE Prophylaxis – Hospital Admission - Documentation of a reason why no mechanical or pharmacological prophylaxis was administered at hospital admission.
Allowable values: Yes or No/UTD.

• Reason for Oral Factor Xa Inhibitor – Documentation of a reason why Oral Factor Xa Inhibitor was administered for VTE prophylaxis.
Allowable values: Yes or No/UTD.

• VTE Prophylaxis – The type of venous thromboembolism prophylaxis documented in the medical record.
Allowable values: 1 Low dose unfractionated heparin (LDUH); 2 Low molecular weight heparin (LMWH); 3 Intermittent pneumatic compression devices (IPC); 4 Graduated compression stockings (GCS); 5 Factor Xa Inhibitor; 6 Warfarin; 7 Venous foot pumps (VFP); 8 Oral Factor Xa Inhibitor; 9 Aspirin; A None of the above or not documented or unable to determine from medical record documentation.

• VTE Prophylaxis Date – The month, day, and year that the initial VTE prophylaxis (mechanical and/or pharmacological) was administered after hospital admission.

Patients are eligible for the numerator population when VTE Prophylaxis equals 1,2,3,5,6,7, or allowable value equals "yes" for Reason for No VTE Prophylaxis-Hospital Admission or "yes" for Reason for Oral Factor Xa Inhibitor and VTE Prophylaxis Date = 0 or 1.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Ischemic or hemorrhagic stroke patients

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Seven data elements are used to calculate the denominator:

1. Admission Date – The month, day and year of admission to acute inpatient care.
2. Birthdate - The month, day and year the patient was born.
3. Clinical Trial - Documentation that during this hospital stay the patient was enrolled in a clinical trial in which patients with stroke were being studied. Allowable values: Yes or No/UTD.
4. Comfort Measures Only – The earliest day the physician/APN/PA documented comfort measures only after hospital arrival.
Allowable values: 1 (Day 0 or 1); 2 (Day 2 or after); 3 (Timing Unclear); 4 (Not Documented/UTD).
5. Discharge Date – The month day and year the patient was discharged from acute care, left against medical advice or expired during the stay.
6. Elective Carotid Intervention – Documentation demonstrates that the current admission is solely for the performance of an elective carotid intervention (e.g., elective carotid endarterectomy, angioplasty, carotid stenting).
Allowable values: Yes or No/UTD.
7. ICD-10-CM Principal Diagnosis Code - The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.

Population: Discharges with ICD-10-CM Principal Diagnosis Code for ischemic or hemorrhagic stroke as defined in Appendix A, Table 8.1 or Table 8.2.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*

- Less than 18 years of age
- Length of Stay < 2 days
- Length of Stay > 120 days
- Comfort measures only documented on day of or day after hospital arrival
- Enrolled in clinical trials related to stroke
- Admitted for elective carotid intervention

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

- The patient age in years is equal to the Discharge Date minus the Birthdate. Patients less than 18 years are excluded.
- The Length of Stay (LOS) in days is equal to the Discharge Date minus the Admission Date. If the LOS is less than 2 days or greater than 120 days, the patient is excluded.
- Patients with Comfort Measures Only allowable value of 1 (Day 0 or 1) are excluded.
- Patients are excluded if "Yes" is selected for Clinical Trial.
- Patients with ICD-10-PCS procedure codes for carotid intervention procedures as identified in Appendix A, Table 8.3, if medical record documentation states that the patient was admitted for the elective performance of this procedure are excluded.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*

Not applicable, the measure is not stratified.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)

No risk adjustment or risk stratification

If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*

Not applicable

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*

Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**

Rate/proportion

If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*

1. Start processing. Run cases that are included in the Stroke (STK) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check Comfort Measures Only

a. If Comfort Measures Only is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.

b. If Comfort Measures Only equals 1, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.

c. If Comfort Measures Only equals 2, 3, or 4, continue processing and proceed to Clinical Trial.

3. Check Clinical Trial

a. If Clinical Trial is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.

b. If Clinical Trial equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Clinical Trial equals No, continue processing and proceed to Elective Carotid Intervention.

4. Check admitted for Elective Carotid Intervention
a. If Elective Carotid Intervention is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Elective Carotid Intervention equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Elective Carotid Intervention equals No, continue processing and proceed to Length of Stay calculation.

5. Calculate the Length of Stay (LOS). Length of Stay, in days, is equal to the Discharge Date minus the Admission Date.

6. Check Length of Stay (LOS)
a. If the Length of Stay is greater than or equal to zero and less than 2, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
b. If the Length of Stay is greater than or equal to 2, continue processing and proceed to VTE Prophylaxis.

7. Check VTE Prophylaxis
a. If VTE Prophylaxis is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If VTE Prophylaxis equals A only, continue processing and proceed to Reason for No VTE Prophylaxis-Hospital Admission.
c. If VTE Prophylaxis equals 1, 2, 3, 4, 5, 6, 7, 8 or 9, continue processing and proceed to step 9 and recheck VTE Prophylaxis.

8. Check Reason for No VTE Prophylaxis-Hospital Admission
a. If Reason for No VTE Prophylaxis-Hospital Admission is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Reason for No VTE Prophylaxis-Hospital Admission equals Yes, the case will proceed to a Measure Category Assignment of E and will be in the Numerator Population. Stop processing.
c. If Reason for No VTE Prophylaxis-Hospital Admission equals No, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.

9. Recheck VTE Prophylaxis
a. If none of the VTE Prophylaxis equals 1, 2, 3, 5, 6 or 7, continue processing and recheck VTE Prophylaxis.
b. If any VTE Prophylaxis equals 1, 2, 3, 5, 6 or 7, continue processing and proceed to step 13 and check VTE Prophylaxis Date.

10. Recheck VTE Prophylaxis
a. If VTE Prophylaxis is not equal to 8, continue processing and proceed to Reasons for No VTE Prophylaxis-Hospital Admission.
b. If any of VTE Prophylaxis equals 8, continue processing and proceed to step 12 and check Reason for Oral Factor Xa Inhibitor.

11. Check Reason for No VTE Prophylaxis-Hospital Admission
a. If Reason for No VTE Prophylaxis-Hospital Admission is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Reason for No VTE Prophylaxis-Hospital Admission equals Yes, the case will proceed to a Measure Category Assignment of E and will be in the Numerator Population. Stop processing.
c. If Reason for No VTE Prophylaxis-Hospital Admission equals No, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.

12. Check Reason for Oral Factor Xa Inhibitor
a. If Reason for Oral Factor Xa Inhibitor is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Reason for Oral Factor Xa Inhibitor equals Yes, continue processing and proceed to VTE Prophylaxis Date.
c. If Reason for Oral Factor Xa Inhibitor equals No, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.

13. Check VTE Prophylaxis Date
a. If VTE Prophylaxis Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.

b. If VTE Prophylaxis Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
c. If the VTE Prophylaxis Date equals a Non-Unable To Determine (non-UTD) Value, continue processing and proceed to VTE Prophylaxis Day calculation.

14. Calculate VTE Prophylaxis Day. The VTE Prophylaxis Day, in days, is equal to the VTE Prophylaxis Date minus the Admission Date.

15. Check VTE Prophylaxis Day
a. If the VTE Prophylaxis Day is equal to zero or 1, the case will proceed to a Measure Category Assignment of E and will be in the Numerator Population. Stop processing.
b. If the VTE Prophylaxis Day is greater than or equal to 2, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
c. If the VTE Prophylaxis Day is less than 0, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter for the measure set cannot sample.

Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

Quarterly Sampling
The Quarterly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing quarterly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If "N" >/= 900, then "n" 180
If "N" 226-899, then "n" 20% of Initial Patient Population size
If "N" 45-225, then "n" 45
If "N" 6-44, No sampling; 100% Initial Patient Population required
If "N" 0-5, Submission of patient level data is not required; if submission occurs, 100% Initial Patient Population required

Monthly Sampling
The Monthly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing monthly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If "N >/= 300, then "n" 60
If "N 76-299, then "n" 20% of Initial Patient Population size
If "N 15-75, then "n" 15
If "N < 15, No sampling; 100% Initial Patient Population required

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the

data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Each data element in the data dictionary includes suggested data sources. The data are collected using contracted Performance Measurement Systems (vendors) that develop data collection tools based on the measure specifications. The tools are verified and tested by Joint Commission staff to confirm the accuracy and conformance of the data collection tool with the measure specifications. The vendor may not offer the measure set to hospitals until verification has been passed.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0434_MeasureTesting_MSF5.0_Data-635905390698284756.doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0434 NQF Project: Neurology Project

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on this measure is as follows:
170 health care organizations representing various hospital demographics:
17 For Profit; 144 Not for Profit; 9 Government
62 >=300 beds; 71 100-299 beds; 37 <100 beds
142 Urban; 28 Rural
22 Teaching; 148 Non-teaching

States represented in this data collection effort include: AL, AR, AZ, CA, CT, FL, GA, IA, ID, IL, IN, KY, LA, MA, MD, MI, MN, MO, MS, NC, NE, NJ, NM, NY, OH, PA, PR, SC, SD, TN, TX, VA, WA, WI

15 performance measurement systems are used for data transmission to The Joint Commission.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with monthly data transmission to The Joint Commission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of approximately 900 patient records identified at 30 pilot sites over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rates of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify potential measure enhancements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures.) The following is a list of the major tests conducted on the submitted data for this measure:

- Transmission of complete data
- Usage of data received (to understand if the healthcare organization provides the relevant service to treat the relevant population)
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements

- Validation of the accuracy of target outliers
- Data integrity
- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors is as follows:
- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Data element agreement rates reported to The Joint Commission for the time period of one year (4Q2010-3Q2011) have shown an overall agreement rate of 97.84%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-1.

Data Elements	Total'n'numerator	Total'n'denominator	Rate
Clinical Trial	712	714	99.7%
Comfort Measures Only	709	714	99.3%
Elective Carotid Intervention	711	714	99.6%
Reason for No VTE Prophylaxis-Hospital Admission	37	45	82.2%
VTE Prophylaxis	293	326	89.9%
VTE Prophylaxis Date	263	281	94.0%

These agreement rates are considered to be well within acceptable levels.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* are consistent with the evidence cited in support of the measure focus (*criterion 1c)* and identify any differences from the evidence:
This measure focuses on ischemic or hemorrhagic stroke patients who received VTE prophylaxis or have documentation why no VTE prophylaxis was given the day of or day after hospital admission. The literature supporting this measure indicates that VTE prophylaxis is appropriate for hospitalized stroke patients because they are at increased risk of developing VTE when compared to other hospitalized medical and surgical patients. This measure excludes patients less than 18 years of age. Also excluded from the measure are patients who have a length of stay (LOS) of less than two days or more than 120 days, those who are enrolled in a clinical trial for stroke or for whom comfort measures only have been ordered on the day of or the day after arrival. These exclusions are not addressed in the literature, but are included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

The measure specifications differ from guideline recommendations by excluding patients admitted for Elective Carotid Intervention. Operationally, patients with a planned admission to the hospital for an elective carotid intervention which restores blood flow to the brain and prevents stroke, such as carotid endarterectomy or carotid stenting, are excluded from the measure.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:
At the time this measure was originally tested measure validity was assessed via survey and focus groups of hospitals participating in the pilot test. All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant.

Since the measure has been in national use, continued validity of the measure has been determined through analysis of feedback from

measure users. The Joint Commission provides a web-based application with which measure users can provide feedback regarding appropriateness of measure specifications, request clarification of specifications, and/or provide other comments pertinent to the measure. This feedback is systematically, continually, reviewed in order to identify trends and to identify areas of the measure specifications that require clarification or revision. Additionally, Joint Commission staff continually monitors the national literature and environment in order to assess continued validity of this measure. An expert technical advisory panel (TAP) meets on a quarterly basis in order to assess continued validity of this measure. And finally, conversion from ICD-9-CM diagnosis codes to ICD-10-CM diagnosis codes has been completed and reviewed by the Technical Advisory Panel for validity. The panel has determined that the intent of the measure has not changed as a result of the conversion. The crosswalk will also be posted in a future version of the specifications manual for public comment, and results of feedback will be reviewed and incorporated into the measure specifications where indicated.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:
Analysis of feedback obtained via our automated feedback system reveals 193 submissions regarding specifications for this measure over the past year. STK-1 receives more questions than other measures in the stroke set, resulting from modification to the measure following CMS alignment with similar Surgical Care Improvement Project (SCIP) VTE and VTE core measures. Questions involve all three numerator data elements VTE Prophylaxis, VTE Prophylaxis Date, and Reason for No VTE Prophylaxis Hospital Admission. Questions about the acceptability of sequential compression devices (SCDs) alone when administered without documentation of a contraindication to pharmacological prophylaxis, as well as the need for VTE prophylaxis for ambulatory patients, are frequently received. The requirement for documentation of both pharmacological and mechanical reasons when neither form of prophylaxis is administered is another common theme. Clarifying notes for abstraction have been added to the data element definition for Reason for No VTE Prophylaxis-Hospital Admission to ease abstraction, as well as, feedback provided to individual users.

Many questions have been received regarding the acceptability of new anticoagulant agents, (i.e., dabigatran, rivaroxaban) for inclusion in the numerator population when these agents are administered the day of or day after hospital admission. Since the dabigatran indication approved by the U.S. Food and Drug (FDA) administration is for prevention of stroke in patients with a diagnosis of atrial fibrillation, but does not include a specific indication for VTE prophylaxis, an abstraction guideline has been added to the data element Reason for No VTE Prophylaxis-Hospital Admission. This abstraction guideline allows patients who receive anticoagulant therapy other than warfarin (e.g., dabigatran) for atrial fibrillation or other conditions the day of or day after hospital admission to be included in the numerator for this measure. The next version of the specifications manual details algorithm revisions to address the appropriate use of rivaroxaban for stroke patients with atrial fibrillation/flutter. In the interim, hospitals have been instructed to apply the same guideline used for dabigatran to include rivaroxaban cases in the numerator.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:
Measure exclusions that were not derived directly from the evidence are presented below. Please note that these are population exclusions that are necessary to ensure consistency in all measures in this 8 measure set.
These exclusions were analyzed for frequency of occurrence. An issue that is of great concern to users of this measure is that due to the presence of exceptions to the measure, attainment of a 100% measure rate is not possible. Because of the role of this measure in the current Joint Commission accreditation process this is especially troubling to measure users. This concern is the basis for a number of the non-evidence-based exclusions to these measures. The following measure exclusions that were not derived directly from the evidence are as follows:

1. Patients < 18 years old
2. Patients who have a Length of Stay (LOS) less than 2 days
3. Patients who have a length of stay (LOS) greater than 120 days
4. Patients with Comfort Measures Only documented on the day of or day after hospital arrival
5. Patients enrolled in clinical trials
6. Patients admitted for Elective Carotid Intervention

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
N=39,812

1. Patients < 18 years old = 0%
2. Patients who have a Length of Stay (LOS) less than 2 days = 1.93%
3. Patients who have a length of stay (LOS) greater than 120 days = 0%
4. Patients with Comfort Measures Only documented on the day of or day after hospital arrival = 0.62%
5. Patients enrolled in clinical trials =0.39%
6. Patients admitted for Elective Carotid Intervention = 9.96%

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Not Applicable

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:
Not Applicable

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:
Not Applicable

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:
Not Applicable

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:
The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
STK-1 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10thPercentile
3Q2011	154	0.88228	0.17825	1	1	0.94666	0.85	0.69231

2Q2011	155	0.88398	0.17796	1	1	0.94595	0.86441	0.67742
1Q2011	160	0.86562	0.17885	1	0.98571	0.92857	0.83485	0.62179
4Q2010	136	0.85448	0.2120	1	0.99738	0.92582	0.83333	0.55319
3Q2010	131	0.82709	0.2078	1	0.97222	0.89744	0.77778	0.57143
2Q2010	122	0.83441	0.1884	1	0.96429	0.89609	0.76471	0.58621
1Q2010	100	0.81077	0.2045	1	0.95676	0.86667	0.75379	0.52273
4Q2009	49	0.76618	0.22024	0.96296	0.92453	0.82813	0.69565	0.46032

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Multiple data sources are not used for this measure.

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:
Not applicable

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:
Not applicable

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: The measure is not stratified for disparities.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:
This measure was originally specified to capture overall rates of VTE prophylaxis for ischemic or hemorrhagic stroke patients with no focus on disparities. The Joint Commission does not currently capture gender-specific data, or data elements for race or ethnicity because these data elements have not been shown to be reliably collectable due to the fact that no national standardized definitions exist for these data elements. Also, not all hospitals collect race and ethnicity. In the future, it may be feasible for The Joint Commission to explore how race and ethnicity and other relevant disparity data, might be collected reliably.

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?
(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

NEW (since 2012 endorsement): Data for Empirical Testing

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

2b2. VALIDITY TESTING
2b2.1. What level of validity testing was conducted? (*may be one or both levels*)
☐ **Critical data elements** (*data element validity must address ALL critical data elements*)
☐ **Performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The data used to measure validity consists of one year of data (third, and fourth quarter 2014 and first and second quarter for 2015) extracted from The Joint Commission warehouse. The hospital selection was based on those hospitals that reported 12 months of data and had 30 or more denominator cases for the year.

Measure convergent and criterion validity was assessed using hospitals patient level data from The Joint Commission warehouse. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous face validity testing. The patient population was comprised of all patients discharged with an ICD-9-CM Principal Diagnosis Code for stroke as defined in Appendix A, Table 8.1 or Table 8.2.

Conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications. The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015.

We conducted a secondary analysis to help interpret results; correlation with other stroke (STK) process measures.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Descriptive statistics for the STK measures:
N=1,318 hospitals
n= 2, 206,379 patients records

STK-1
Median: 98%
Percentile 10%: 93%
Percentile 25%: 97%
Percentile 75%: 100%
Percentile 90%: 100%

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations							
	STK_1	STK_2	STK_3	STK_4	STK_5	STK_6	STK_10
STK_1	1.00000 1318	0.55423 <.0001 1318	0.30311 <.0001 1302	0.37009 <.0001 1144	0.52785 <.0001 1318	0.53082 <.0001 1318	0.45291 <.0001 1318
STK_2	0.55423 <.0001 1318	1.00000 1318	0.37566 <.0001 1302	0.28169 <.0001 1144	0.53355 <.0001 1318	0.47326 <.0001 1318	0.29578 <.0001 1318
STK_3	0.30311 <.0001 1302	0.37566 <.0001 1302	1.00000 1302	0.25551 <.0001 1137	0.25665 <.0001 1302	0.28198 <.0001 1302	0.15819 <.0001 1302
STK_4	0.37009 <.0001 1144	0.28169 <.0001 1144	0.25551 <.0001 1137	1.00000 1144	0.22516 <.0001 1144	0.43717 <.0001 1144	0.41076 <.0001 1144
STK_5	0.52785 <.0001 1318	0.53355 <.0001 1318	0.25665 <.0001 1302	0.22516 <.0001 1144	1.00000 1318	0.35302 <.0001 1318	0.39079 <.0001 1318
STK_6	0.53082 <.0001 1318	0.47326 <.0001 1318	0.28198 <.0001 1302	0.43717 <.0001 1144	0.35302 <.0001 1318	1.00000 1318	0.45420 <.0001 1318
STK_10	0.45291 <.0001 1318	0.29578 <.0001 1318	0.15819 <.0001 1302	0.41076 <.0001 1144	0.39079 <.0001 1318	0.45420 <.0001 1318	1.00000 1318

The correlations of the stroke measures are all positively correlated and statistically significant indicating that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)
Overall, the positive inter-correlations indicate convergent validity of the measures.
STK-01 is positively correlated with other stroke evidence-based processes of care.

2b3. EXCLUSIONS ANALYSIS .
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patients who have a Length of Stay less than 2 days
- Patient who have a Length of Stay greater than 120 days

- Patients with *Comfort Measures* documented on day of or day after arrival
- Patient enrolled in a Clinical Trial
- Patients admitted for *Elective Carotid Intervention*

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are detailed below:

Exclusion: *Comfort Measures Only* documented on day of or day after arrival
Overall Occurrence n = 28,126
Overall Occurrence Percentage: 4.46%
Minimum: 0.31%
10th Percentile: 1.67%
Median: 4.40%
90th Percentile: 9.09%
Maximum: 30.8%

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%
10th Percentile: 0.24%
Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for *Elective Carotid Intervention*
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.6%
90th Percentile: 25.31%
Maximum: 95.3%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The median frequency of exclusions range from low to moderate. The distribution of exclusions across hospitals is generally narrow indicating that the occurrence is random and likely would not bias performance results, although the percentage of patients excluded may differ depending on whether the hospital was a stroke center.

For criterion validity, it is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a Length of Stay greater than 120 days

Rationale: Included to harmonize with other CMS/Joint Commission aligned measures.

Patients who have a Length of Stay less than 2
Rationale: Inclusion of patients with a Length of Stay shorter than 2 days may falsely increase the denominator population.

Patients with *Comfort Measures Only* documented on day of or day after arrival
Rationale: It is inappropriate to treat such patients beyond those measures necessary for comfort.

Patients enrolled in a Clinical Trial
Rationale: *VTE Prophylaxis* as specified by this measure may compromise a clinical trial.

Patients admitted for *Elective Carotid Intervention*
Rationale: Patients who are not admitted for treatment of an acute stroke may not require VTE prophylaxis.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other

If other: Data element allowable values are selected, either manually or electronically, from clinical and coded data available in medical record documentation. All medical record documentation is used in the abstraction process. Vendor data collection tools are used to import data elements needed for measure rate calculation.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract the electronic version of this measure, so continues to offer this chart-abstracted version which allows for data capture from unstructured data fields.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

At the present time, hospitals using this performance measure generally collect measure data via manual review of the paper medical record, the EHR or a combination of both. Collected data are submitted to The Joint Commission on a quarterly basis, by way of contracted performance measurement system vendors, as described previously. Specifications for this measure are freely available to anyone who wishes to use the measure. Feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As described above, as feedback from measure users has indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

There are no fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Quality Check® http://www.qualitycheck.org/consumer/searchQCR.aspx Hospital Compare https://www.medicare.gov/hospitalcompare/search.html Public Health/Disease Surveillance Paul Coverdell National Acute Stroke Registry http://www.cdc.gov/dhdsp/programs/stroke_registry.htm Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program XX Hospital Accreditation Program http://www.jointcommission.org/ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Annual Report – Improving America's Hospitals http://www.jointcommission.org/annualreport.aspx Quality Improvement (Internal to the specific organization) Disease-Specific Care Certification for Comprehensive Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx Disease-Specific Care Certification for Primary Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

• Name of program and sponsor: Quality Check®; The Joint Commission

• Purpose: A public website that allows consumers to: search for accredited and certified organizations by city and state, by name or by zip code (up to 250 miles); find organizations by type of service provided within a geographic area; download free hospital performance measure results; and, print a list of Joint Commission certified disease-specific care programs and health care staffing firms.

• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

• Name of program and sponsor: Hospital Compare; Centers for Medicare & Medicaid Services
• Purpose: A public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4000+ Medicare-certified hospitals (2015)

• Name of program and sponsor: Paul Coverdell National Acute Stroke Registry; Centers for Disease Control and Prevention
• Purpose: A national registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.
• Geographic area and number and percentage of accountable entities and patients included: 11 states; 403 hospitals (CDC, 2014)

• Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
• Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 hospitals (2015)

• Name of program and sponsor: Annual Report-Improving America's Hospitals; The Joint Commission
• Purpose: The annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

• Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

• Name of program and sponsor Disease-Specific Care Certification for Primary Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
• Geographic area and number and percentage of accountable entities and patients included Nationwide; 1079 hospitals

• Name of program and sponsor Hospital Accreditation Program; The Joint Commission
• Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
• Geographic area and number and percentage of accountable entities and patients included Nationwide; 3300 Joint Commission-accredited hospitals (2014)

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in

use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Progress (trends in performance results, number and percentage of people receiving high-quality healthcare:
The rate of VTE prophylaxis has steadily increased over the past five years based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 97.3%. A gap of 9-10% still exists for the 10th percentile of hospitals. Modest differences in VTE prophylaxis in black patients have been reported by GWTG (Schwamm, 2010).

Geographic area and number and percentage of accountable entities and patients included:
Nationwide; 1299 hospitals; 213,000 patients (The Joint Commission, 2014)

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
As a requirement for inclusion of the stroke measures as a core measure set in the Specifications Manual for National Hospital Inpatient Quality Measures (October 2009), the STK-1 algorithm was revised to align with specifications for similar core measure sets. Subsequently, allowable value (9) aspirin was added to the data element VTE Prophylaxis for the Surgical Care Improvement Project (SCIP) VTE measures. The addition of aspirin in the data element caused a need to educate stroke abstractors that aspirin alone without another form of VTE Prophylaxis or a documented reason why only aspirin was administered for prophylaxis is needed, as aspirin is not approved by the U.S. Food and Drug Administration (FDA) for VTE prophylaxis in the stroke population.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0218 : Surgery Patients Who Received Appropriate Venous Thromboembolism Prophylaxis Within 24 Hours Prior to Surgery to 24 Hours After Surgery
0239 : Perioperative Care: Venous Thromboembolism (VTE) Prophylaxis
0371 : Venous Thromboembolism Prophylaxis
0372 : Intensive Care Unit Venous Thromboembolism Prophylaxis

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0217 : Surgery Patients with Recommended Venous Thromboembolism (VTE) Prophylaxis Ordered; Centers for Medicare and Medicaid Services

0240: Stroke and Stroke Rehabilitation: Venous Thromboembolism (VTE) Prophylaxis for Ischemic Stroke or Intracranial Hemorrhage; American Academy of Neurology

5a. Harmonization

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Measures NQF# 0371 and NQF# 0372 are Venous Thromboembolism (VTE) measures which specifically exclude the stroke population. The measures are completely harmonized in terms of measure specifications and data element definitions; NQF# 0218 addresses the surgical population only, and therefore do not apply to stroke patients. Common data elements with this measure have been completely harmonized. Measure 0239 is a physician performance measure with a targeted population of surgical patients identified through CPT codes and thus is a different level of measurement. This measure evaluates physician practice as opposed to hospital processes.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not Applicable

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0435
Measure Title: STK 02: Discharged on Antithrombotic Therapy
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of ischemic stroke patients prescribed antithrombotic therapy at hospital discharge.

This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-4: Thrombolytic Therapy,STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-6 Discharged on Statin Medication, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs.
Developer Rationale: Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. The effectiveness of antithrombotic agents in reducing stroke mortality, stroke-related morbidity and recurrence rates has been studied in several large clinical trials. Studies have shown that antithrombotic therapy prescribed at discharge following acute ischemic stroke reduces stroke mortality and morbidity.

Healthcare organizations that track antithrombotic therapy prescribed at discharge for internal quality improvement purposes have seen an improvement in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

Numerator Statement: Ischemic stroke patients prescribed antithrombotic therapy at hospital discharge
Denominator Statement: Ischemic stroke patients
Denominator Exclusions:
- Less than 18 years of age
- Length of Stay > 120 days
- Comfort measures only documented
- Enrolled in clinical trials related to stroke
- Admitted for elective carotid intervention
- Discharged to another hospital
- Left against medical advice
- Expired
- Discharged to home for hospice care
- Discharged to a health care facility for hospice care
- Documented reason for not prescribing antithrombotic therapy at discharge

Measure Type: Process
Data Source: Electronic Clinical Data, Paper Medical Records

Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2008 **Most Recent Endorsement Date:** Nov 01, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Summary of prior review in 2012:

- The body of evidence consistently supports that prolonged antithrombotic therapy reduces the risk of death in patients with a prior history of ischemic stroke. Based on these findings, antithrombotic therapy should be prescribed at discharge following acute ischemic stroke, unless there are contraindications to therapy.
- 2010 AHA/ASA Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack, page 249.
    - Class I, Level of Evidence A - For patients with noncardioembolic ischemic stroke or TIA, the use of antiplatelet agents rather than oral anticoagulation is recommended to reduce the risk of recurrent stroke or another cardiovascular event.
    - Aspirin (50 mg/d to 325 mg/d) monotherapy (Class I, Level of Evidence A), the combination of aspirin 25 mg and extended-release dipyridamole 200 mg twice daily (Class I, Level of Evidence B), and clopidogrel 75 mg monotherapy daily (Class IIa, Level of Evidence B), are all acceptable options for initial therapy. The selection of an antiplatelet agent should be individualized on the basis of patient risk factor profiles, cost, tolerance, and other clinical characteristics.
- The 2012 Committee expressed no concerns regarding the evidence underlying this measure.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☐ **The developer provided updated evidence for this measure:**

Guidance from the Evidence Algorithm:

Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency high(Box 5) → Box 5a High

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and voting on Evidence?*

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer provides the following national trend data:

	CY 2010	CY2011	CY 2012	CY 2013	CY 2014
# hospitals	137	157	158	262	1299
# cases (den)	19,889	24,053	24,420	37,661	180,048
National aggregate rate	0.99	0.99	0.99	0.99	0.99
Mean hospital rate	0.98	0.97	0.99	0.99	0.99
50th percentile	1.00	1.00	1.00	1.00	1.00
10th and 90th percentiles	0.94 (10th) 1.00 (90th)	0.95 (10th) 1.00 (90th)	0.97 (10th) 1.00 (90th)	0.98 (10th) 1.00 (90th)	0.98 (10th) 1.00 (90th)

- Participation in this measure has grown significantly in the past five years. National hospital performance seems to be very high with little to no room for improvement.
- In 2012, the Committee noted the overall high rate of performance for this measure (approximately 98% among reporting hospitals).

Disparities:

- The developer does not provide disparities data from use of this measure.
- Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age."

Questions for the Committee:

- *Data for 2014 includes 1,299 hospitals. What proportion of hospitals caring for stroke patients is captured in this data?*
- *Performance over five years has been consistently high. How much further improvement in performance is likely using this measure?*
- *Does a gap in care still exist that warrants this national performance measure?*
- *How can this measure be used to better understand disparities in care and outcomes for certain groups?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **Not aware of any new studies/information

**The measure documents the proportion of patients with ischemic stroke prescribed antithrombotic therapy at hospital discharge. There is ample empirical evidence supporting the benefits antithrombotic therapy at discharge for secondary prevention of subsequent stroke or other cardiovascular events as well as slight but significant decrease in mortality.

**I am not aware of any new information that changes the evidence for this measure.

1b. Performance Gap

Comments: **No disparities data provided

**Participation in this measure has increased over time Evidence supports an increase in national hospital performance -with an approximate rate of 98% for all reporting hospitals in 2014-data also suggest little to no improvement . However, substantial racial/ethnic disparities in stroke mortality and morbidity are well documented. Despite evidence of overall increased use of antithrobotics at discharge -a 2013 study(Quian et al) as well as CDC 2014 reports indicate that disparities still exist.

**There was performance gap data. However, it showed that the gap has narrowed substantially since the initiation of this measure. Compliance with the standard is now so high that those cases in which included patients were not discharged on antithrombotic therapy may well have been failure to recognize an indication in the chart as to the reason, or valid reasons not included in the specification. There is little performance gap remaining to close.

1c. High Priority (previously referred to as High Impact)

Comments: **NA

**Yes

**Yes. stated and logical

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): medical record abstraction (paper or electronic). This is not an eMeasure.

Specifications:

- One data element defines the numerator and nine data elements define the denominator.
- Developer reports that "Since the last endorsement date, the measure rationale for STK-02: Discharged on Antithrombotic Therapy was updated to address the use of novel oral anticoagulant drugs in stroke patients requiring antithrombotic therapy. In recent years, direct oral anticoagulant agents (DOACs) [A.K.A. novel oral anticoagulant agents (NOACs)] have been developed and approved by the U.S. Food and Drug Administration (FDA) for stroke prevention, and may be considered as an alternative to warfarin for select patients. Following FDA new drug approval of these agents, several DOACs were added to the medication table for antithrombotic medications (Appendix C, Table 8.2) used for abstraction of the measure."
- Diagnosis and procedure codes have been converted from ICD-9 to ICD-10-CM diagnosis and ICD-10-PCS procedure codes. See attached spreadsheet Appendix A Table 8.1 (ischemic stroke).
- A conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications:
    - "The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015 developer advisory panel determined that the intent of the measure was changed by the conversion to ICD-10 and a crosswalk was posted."
- Sampling monthly or quarterly is allowed; instructions for sampling are provided.

- A web-based data collection tool has been developed but is not described. Hospitals are not required to use this tool.
- The measures is specified at the hospital level of analysis.

Questions for the Committee :

- *Are changes to the measure specifications appropriate?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Inter rater-reliability of the data elements for one year (4Q2010-3Q2011) in 77 hospitals and 739 patient records showed an overall agreement rate of 97.61%.
  - One data element found less than 95% agreement: Reason for Not Prescribing Antithrombotic Therapy at Discharge (91%). No Kappa scores were presented.

Describe any updates to testing: No new information provided

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing: see above

Results of reliability testing: see above

Guidance from the Reliability Algorithm:

Precise specifications (Box 1) → empirical testing as specified (Box 2) → reliability testing with patient-level data elements (Box 8) → assessed all critical data elements (Box 9) → high/moderate certainty the data elements are reliable (Box 10) → Moderate (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Questions for the Committee:

- *The developer has not provided any new reliability testing. Does the Committee agree there is no need for repeat discussion and voting on Reliability?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Face validity of the data was assessed via survey and focus groups of hospitals participating in the pilot test.
- The developer reports that "All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant."
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.
- Ongoing feedback from measure users that is systematically reviewed to identify trends and/or measure specifications that require clarification or revision. Predominantly, questions involve the data elements Reason for Not Prescribing Antithrombotic Therapy at Discharge. In addition, developers received questions regarding acceptability of new anticoagulant agents.

Describe any updates to validity testing – New empirical validity testing data is presented

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- Pearson Correlation Coefficients were calculated for seven stroke measures for hospitals that reported 12 months of data and had 30 or more denominator cases for the year (2014-2015).Presumably high performing hospitals would do well on all stroke performance measures.

Validity testing results:

- Analysis of 1,318 hospitals and 2,206,379 patients records generated a table of Pearson Correlation Coefficient results that shows a statistically significant (P<.0001), positive correlation of this measure with six other stroke performance measures supporting the hypothesis that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients are excluded from the measure for the following reasons:
  - Patients less than 18 years of age
  - Patient who have a Length of Stay greater than 120 days
  - Patients with Comfort Measures

- Patient enrolled in a Clinical Trial
- Patients admitted for Elective Carotid Intervention
- Patients with a documented Reason For Not Prescribing Antithrombotic Therapy at Discharge
- Patients discharged to another hospital
- Patients who left against medical advice
- Patients who expired
- Patients discharged to home for hospice care
- Patients discharged to a health care facility for hospice care

- New data on the frequency of exclusions is presented for 2, 206,379 admissions in 1,318 hospitals:
  - Comfort Measures Only: 10.4% (range 0.30-35.7%)
  - Clinical Trial: 0.25% (range 0.08-35.5%)
  - Patients admitted for Elective Carotid Intervention: 11.8% (range 0.16-95.2%)
  - Patients with a documented Reason For Not Prescribing Antithrombotic Therapy at Discharge: 6.27% (range 0.27%-60.6%)
  - Patients discharged to another hospital: 35.7% (range 0.79-76.2%)
  - Patients who left against medical advice: 0.67% (range 0.06-9.67%)
  - Patients who expired: 6.04% (range 0.39-20.1%)
  - Patients discharged to home for hospice care: 52.2% (range 6.25-94.3%)
  - Patients discharged to a health care facility for hospice care: 3% (range 0.16-23%)

Questions for the Committee:

- *In this sample approximately half of the patients are excluded. Does the large number of exclusions for Patients discharged to home for hospice care pose a threat to validity of the measure?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Performance data is presented above under opportunity for improvement. Complete details of the data is presented in 1b.
- The developer explains their approach: "There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations."

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods: Not Applicable

2b7. Missing Data

- The developer describes the handling of missing data in the measure specifications (S.22) but does not provide information on the frequency of missing data or potential impact on measure results.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → potential threats to validity mostly assessed (Box2) →empirical validity testing (Box 3) → measure score testing (Box 6) → sound conceptualization and hypotheses (Box 7) →moderate confidence that score are a valid indicator of quality(Box 8a) → High

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **Specifications consistent with evidence

**none

**Specifications are consistent with the evidence

2a2. Reliability Testing

Comments: **New validity testing had adequate scope and method

**yes

**Validity was tested with focus groups. Demonstrated that performance on this measure was highly correlated with those of 7 stroke measures.

2b2. Validity Testing

Comments: **No

**Possibly-data are based on chart extraction for some hospitals -not all extract data from EHR. Given the large volume of stroke patients some error in data collection expected. Closer examination of patients excluded is warrented- particular attention to racial/ethnic characteristics of excluded patients would be useful.

**Exclusions are consistent with the evidence. The large number of exclusions is important given the demographics of the stroke population- many patients are discharged to hospice care.

Risk adjustment not used

Meaningful differences. The measure performance is meaningful. However, compliance for included patients is so high that there are few differences to detect.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **No new reliability data provided

**yes

**Reliability testing was adequate

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement. [feasibility]

- Data source is medical record abstraction: not all hospitals are able to generate the data electronically so a paper-based option is available
- Some data elements are in electronic form
- Data collection burden is significant.

Questions for the Committee:

- o *Are the required data elements routinely generated and used during care delivery?*
- o *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- o *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments: **No concerns
**Medical record abstraction in the absence of EHR data is burdensome.
**All data elements are routinely generated and readily available in electronic form. Data collection may be tedious in the absence of an electronic health record.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- The Joint Commission Quality Check®: Public website that allows consumers to search for accredited and certified organizations.
- CMS Hospital Compare: Public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.
- CDC Paul Coverdell National Acute Stroke Registry: National registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.
- CMS Hospital Inpatient Quality Reporting Program (IQR): The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- The Joint Commission Annual Report-Improving America's Hospitals: Annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
- The Joint Commission Disease-Specific Care Certification for Comprehensive Stroke Centers: Certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
- The Joint Commission Disease-Specific Care Certification for Primary Stroke Centers: Certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results:

- The developer provided the following progress information on 1,256 hospitals nationwide and 180,048 patients:
    - The rate of antithrombotic therapy prescribed at discharge has steadily increased over the past five

years based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 99.4%. A modest gap of approximately 3% still exists for the 10th percentile of hospitals. This trend is consistent with increases in the rate of antithrombotic therapy at discharge published by GWTG-Stroke and PCNASR.
    - Compared to antithrombotic at discharge rates of approximately 75% at the turn of the millennium (Lichtman JH, 2011), the performance gap has significantly narrowed.
    - Significant and important differences between races in rates of antithrombotic therapy prescribed at discharge persist (Schwamm, 2010, Qian, 3013; CDC 2014).

Unexpected findings (positive or negative) during implementation:

- Developer did not identify unexpected findings during implementation.

Potential harms:

- Developer did not identify any unintended consequences related to this measure.

Feedback :

- The developer states that feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As previously noted, measure users have indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences
Comments: **Agree high usability
**The reporting is very straightforward and easy to interpret. Given the well established utility of antithrobotic therapy and the low cost of aspirin -the benefits of this measure outweigh the burden of data collection.
**The measure is reported by the Joint Commission, CDC and CMS

Criterion 5: Related and Competing Measures

Related or competing measures:

- 0068 : Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet
- 0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy
- 0438 : STK 05: Antithrombotic Therapy By End of Hospital Day Two
- 0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy – American Academy of Neurology – not NQF endorsed

Harmonization :

- In the prior evaluation of #0325 and #0435, the Committee expressed a desire for a single measure that could be used at the clinician and facility levels. The two developers agreed it was not feasible at that time to create a single measure. The Committee recommended continued and aggressive efforts for harmonization when possible, and requested an update on progress on harmonization at the time of annual review.
- In 2012 the Committee suggested that the developer consider developing a composite measure that included

#0435 and #0438; they noted that such a composite measure would indicate the percentage of patients who receive appropriate care at both time points and that such a measure would likely provide more opportunity for improvement. The Committee asked the developer if they could provide data on the percentage of patients who get antithrombotic therapy at discharge but not on day two and vice-versa.

- The Joint Commission conducted further data analysis of these measures. Data were collected for the period 4Q2010 to 3Q2011. During this one-year period, data from 39,812 patient records were collected for all stroke (STK) measures. Altogether there were 15,789 patients that received antithrombotic therapy at discharge or by the end of hospital day 2 and were included in the measure population of both measures. Of these, 97 (0.6%) did not have therapy prescribed at discharge and 267 (1.7%) did not have therapy on day two.

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0435 NQF Project: Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The focus of the measure is to prevent a secondary stroke following a new ischemic stroke. Antithrombotic prescribed at discharge >> secondary stroke prevention >>decreased morbidity and mortality.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
This measure is consistent with the guidelines recommended by the American Heart Association/American Stroke Association for the prevention of stroke in patients with stroke or transient ischemic attack. The Antiplatelet Trialists' Collaboration (1994) and later Antithrombotic Trialists' Collaboration (2002) have confirmed that the benefit of long-term antithrombotic therapy to patients with ischemic stroke is well established.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: In 2002, the Antithrombotic Trialists' Collaboration published a systematic overview of the relevant literature supporting antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. The collaboration reviewed 287 studies involving 135,000 patients in comparisons of antiplatelet therapy versus control and 77,000 patients in comparisons of different antiplatelet regimens. Twenty-one of the 287 trials were randomized control trials (n=18,270) involving patients with a history of stroke or transient ischemic attack allocated to a mean duration of 29 months of antiplatelet therapy. Since the previous meta-analysis (1994; 10,255 patients in 18 trials), the amount of information available on the effects of long-term antithrombotic therapy among patients with a history of stroke or transient ischemic attack has substantially increased, primarily due to the results of the second European Stroke Prevention Study (n-6602).

Relevant trials were identified through electronic database searches (MEDLINE, Embase, Derwent, Scisearch, and Biosis), searching the trials registers of the Cochrane Stroke and Peripheral Vascular Disease Groups; manual searching of journals, abstracts, and proceedings of meetings; scrutinizing the reference lists of trials and review articles; and professional inquiry, including colleagues and representatives of pharmaceutical companies.

Trials available by September 1997 that compared an antiplatelet regimen with a control or one antiplatelet regimen with another among patients considered to be high annual risk (> 3%/year) of vascular events because of evidence of pre-existing disease were included in the meta-analysis. Only those trials believed to have used a randomization method that precluded prior knowledge of the next treatment allocated and contained two randomized groups that differed only with respect to the antiplatet comparison of interest were selected. Furthermore, trials of oral antiplatet regimens were eligible only if they had assessed more than one day of treatment. An antiplatelet drug was defined as one whose primary effect on the vascular system inhibits platelet adhesion, platelet aggregation, or both. Details about the method of randomization, blinding of treatment allocation, scheduled duration of treatment, and, if, different,

scheduled duration of follow-up were requested of all trial coordinators.

The primary measure of outcome was a "serious vascular event" (i.e., non-fatal myocardial infarction, non-fatal stroke, or death from a vascular cause including deaths from an unknown cause). Deaths were divided into those with a vascular or non-vascular cause. Strokes were subdivided into intracranial hemorrhages (including intracerebral, subdural, subarachnoid, and extradural hemorrhages), ischemic strokes, and strokes of unknown etiology.

In addition to the meta-analysis from the Antithrombotic Trialists' Collaboration, the following trials relevant to antithrombotic therapy for noncardioembolic stroke were noted in the literature. The landmark trials primarily focused on aspirin and included: the International Stroke Trial Collaborative Group (1994); the Chinese Acute Stroke Trial Collaborative Group (1997); the European Stroke Prevention Study Group 2 (1997); the Canadian Cooperative Study Group (1978); the UK-TIA Study Group (1991); the Dutch TIA Study Group (1991); and the SALT Collaborative Group (1991). Three RCTs evaluated ticlopidine: the Canadian American Ticlopidine Study (1989), the Ticlopidine Aspirin Stroke Study (1989), and the African American Antiplatelet Stroke Prevention Study (2003). Clopidogrel was compared to aspirin alone in the Clopidogrel versus Aspirin in Patients at Risk of Ischemic Events (CAPRIE) trial (1996). The Management of Atherothrombosis with Clopidogrel in High-Risk Patients with Transient Ischemic Attacks or Ischemic Strokes (MATCH) trial (2004) evaluated the effectiveness of clopidogrel 75mg and aspirin 75 mg, compared with clopidogrel 75mg alone for stroke prevention. Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization, Management, and Avoidance (CHARISMA) trial (2006) and the Fast Assessment of Stroke and Transient ischemic attack to prevent Early Recurrence (FASTER) trial (2007) evaluated clopigogrel and aspirin in combination with aspirin alone. Four large RCTs evaluated the effectiveness of dipyridamole and aspirin among patients with stroke or transient ischemic attack: European Stroke Prevention Studies (ESPS-1 1987 and ESPS-2 1997); the European Australian Stroke Prevention in Reversible Ischemic Stroke (ESPRIT) trial (2006); and the Prevention Regimen for Effectively Avoiding Second Strokes (PRoFESS) trial (2008).

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting antithrombotic therapy for secondary stroke prevention is high. Antithrombotic therapy post ischemic stroke is protective for high risk patients, such as, those with a history of prior stroke or vascular event. The body of evidence established by the Antithrombotic Trialists' Collaboration confirmed that prolonged antithrombotic therapy reduces the relative risk of stroke, myocardial infarction or death by approximately 22%. Risk of gastrointestinal hemorrhage or other major hemorrhage is the primary concern with ongoing therapy, although the studies have demonstrated that the risk is low and absolute benefits significantly outweigh the risks.

There was no appreciable evidence that either higher aspirin dose or any other antiplatelet regimen was more effective than low dose aspirin (75 mg to 150 mg) in preventing vascular events. Additionally, the body of evidence did not establish the optimal duration of treatment. The majority of studies reviewed lasted one to three years. The evidence seems to support that longer treatment might be more effective due to significant further benefit noted ($P<0.00001$) between year one and year three for most studies. Adding a second antithrombotic to aspirin may produce additional benefit; however, the evidence is inconclusive in this respect and more research recommended.

In reviewing the literature, there have been more studies conducted than included in the Antithombotic Trialists' meta-analysis. Some studies were not included in the meta-analysis due to the method of randomization chosen or small sample sizes. Antithrombotic medications other than aspirin have not been studied as extensively as aspirin.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of evidence consistently supports that prolonged antithrombotic therapy reduces the risk of death in patients with a prior history of ischemic stroke. Based on these findings, antithrombotic therapy should be prescribed at discharge following acute ischemic stroke, unless there are contraindications to therapy.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
There is strong evidence on the immediate hazards and net benefits of long-term antithrombotic therapy for the prevention of stroke recurrence. The Antithrombotic Trialists' Collaboration meta-analysis found a marginally significant ($P=0.04$) reduction in vascular mortality (7 deaths per 1,000 patients with a history of stroke or transient ischemic attack treated with antithrombotic therapy); however, the reduction in non-fatal vascular events and in all cause mortality was highly significant ($P=0.002$; 15 deaths per 1,000 patients

treated). These benefits exceeded the risk of bleeding (1-2 additional extracranial bleeds per 1,000 patients/year). On average, antithrombotic medications reduce the relative risk of stroke, MI, or death, but important differences exist between the various drugs in this category which directly impacts therapeutic selection.

Considering the cost of aspirin compared to the mean lifetime cost of ischemic stroke, estimated at $140,048 per person, antithrombotic therapy as a secondary stroke prevention strategy is clearly cost-effective. Gaspoz and colleagues (2002) concluded that the extension of aspirin therapy from the current levels of use to all eligible patients for 25 years would have an estimated cost-effectiveness ratio of about $11,000 per quality-adjusted year of life gained. The addition of clopidogrel for the 5 percent of patients who are ineligible for aspirin would cost about $31,000 per quality-adjusted year of life gained. Clopidogrel alone in all patients or in routine combination with aspirin had an incremental cost of more than $130,000 per quality-adjusted year of life gained and remained financially unattractive across a wide range of assumptions. However, clopidogrel alone or in combination with aspirin would cost less than $50,000 per quality-adjusted year of life gained if its price were reduced by 70 to 82 percent, to $1.00 and $0.60 per day, respectively.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: American Heart Association. During our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, and provided information that was accessible and met the requirements set out in this measure maintenance form.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.13 Grade Assigned to the Body of Evidence: Class I, Level of Evidence A

1c.14 Summary of Controversy/Contradictory Evidence: The benefit of prolonged antithrombotic therapy post ischemic stroke is undisputed. No position against long-term antithrombotic therapy was noted in the literature. However, there is still not enough evidence to answer some questions. More randomized trials are needed to determine the optimal therapeutic agent(s) and regimen(s), as well as, the optimal duration of therapy.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and thrombolytic therapy for ischemic stroke: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest. 2004; 126;483S-512S.
- Antiplatelet Trialists' Collaboration. Collaborative overview of randomised trials of antiplatelet therapy, I: prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. BMJ. 1994;308:81–106.
- Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomized trials of anti-platelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients [published correction appears in BMJ. 2002;324:141]. BMJ. 2002;324:71– 86.
- Bhatt DL, Fox KAA, Hacke W, Berger PA, Black HR, Boden WE, Cacoub P, Cohen EA, Creager MA, Easton J, Flather M, Haffner S, Hamm C, Hankey G, Johnston S, Mak K, Mas J, Montalescot G, Pearson T, Steg P, Steinhubl S, Weber M, Brennan D, Fabry-Ribaudo L, Booth J, Topol E; CHARISMA investigators. Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. NEJM. 2006;354:1706 –1717.
- CAPRIE Steering Committee. A randomized, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events (CAPRIE). Lancet. 1996;348:1329 –1339.
- Diener H-C, Bogousslavsky J, Brass LM, Cimminiello C, Csiba L, Kaste M, Leys D, Matias-Guiu J, Rupprecht H-J; on behalf of the MATCH investigators. Aspirin and clopidogrel compared with clopidogrel alone after ischaemic stroke or transient ischaemic attack in high-risk patients (MATCH): randomized, double-blind, placebo-controlled trial. Lancet.

2004;364:331–337.
• Gaspoz J, Coxson PG, Goldman PA, Williams LW, Kuntz KM, Hunink MGM, Goldman L. Cost effectiveness of aspirin, clopidogrel, or both for secondary prevention of coronary heart disease. NEJM. 2002;346:1800-1806.
• Gorelick PB, Richardson D, Kelly M, Ruland S, Hung E, Harris Y, Kittner S, Leurgans S; for the African American Antiplatelet Stroke Prevention Study (AAASPS) Investigators. Aspirin and ticlopidine for prevention of recurrent stroke in black patients. JAMA. 2003;289: 2947–2957.
• Gent M, Easton JD, Hachinski VC, Panak E, Sicurella J, Blakely JA, Ellis DJ, Harbison JW, Roberts RS, Turpie AGG. The Canadian American Ticlopidine Study (CATS) in Thromboembolic Stroke. Lancet. 1989;1215-1220.
• Guyatt GH, Akl EA, Crowther M, Gutterman DD, Sch?nemann HJ, and for the American College of Chest Physicians Antithrombotic Therapy and Prevention of Thrombosis Panel. Executive summary: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141: 34S.
• Guyatt G, Schunëmann H, Cook D, Jaeschke R, Pauker S, Bucher H. Grades of recommendation for antithrombotic agents. Chest. 2001;119:3S-7S.
• Hass WK, Easton JD, Adams HP, Pryse-Phillips W, Molony BA, Anderson S, Kamm B; for the Ticlopidine Aspirin Stroke Study Group. A randomized trial comparing ticlopidine hydrochloride with aspirin for the prevention of stroke in high-risk patients. N Engl J Med. 1989;321: 501–507.
• Johnson ES, Lanes SF, Wentworth CE, Satterfield MH, Abebe BL, Dicker LW. A meta-regression analysis of the dose-response effect of aspirin on stroke. Arch Intern Med. 1999;159:1248 –1253.
• Kennedy J, Hill MD, Ryckborst KJ, Eliasziw M, Demchuk AM, Buchan AM. Fast assessment of stroke and transient ischaemic attack to prevent early recurrence (FASTER): a randomised controlled pilot trial. Lancet Neurol. 2007;6:961–969.
• Sacco RL, Diener H-C, Yusuf S, Cotton D, Ounpuu S, Lawton WA, Palesch Y, Martin RH, Albers GW, Bath P, Bornstein N, Chan BP, Chen ST, Cunha L, Dahlo¨f B, De Keyser J, Donnan GA, Estol C, Gorelick P, Gu V, Hermansson K, Hilbrich L, Kaste M, Lu C, Machnig T, Pais P, Roberts R, Skvortsova V, Teal P, Toni D, Vandermaelen C, Voigt T, Weber M, Yoon BW; PRoFESS Study Group. Aspirin and extended release dipyridamole versus clopidogrel for recurrent stroke. N Engl J Med. 2008;359:1238 –1251.
• Sandercock P, Collins R, Counsell C, Farrell B, Peto R, Slattery J, Warlow C, International Stroke Trial Collaborative Group: The International Stroke Trial (IST): A randomised trial of aspirin, subcutaneous heparin, both, or neither among 19,435 patients with acute ischaemic stroke. Lancet. 1997;349:1569-1581.
• The Canadian Cooperative Study Group. A randomized trial of aspirin and sulfinpyrazone in threatened stroke. N Engl J Med. 1978;299:53–59.
• The Dutch TIA Trial Study Group. A comparison of two doses of aspirin (30 mg vs. 283 mg a day) in patients after a transient ischemic attack or minor ischemic stroke. N Engl J Med. 1991;325:1261–1266.
• The ESPRIT Study Group. Aspirin plus dipyridamole versus aspirin alone after cerebral ischaemia of arterial origin (ESPRIT): randomised controlled trial. Lancet. 2006;367:1665–1673.
• The ESPS Group. The European Stroke Prevention Study (ESPS): principal end-points. Lancet. 1987:2:1351–1354.
• The SALT Collaborative Group. Swedish Aspirin Low-dose Trial (SALT) of 75 mg aspirin as secondary prophylaxis after cerebrovascular ischaemic events. Lancet. 1991;338:1345–1349.
• UK-TIA Study Group. The United Kingdom Transient Ischaemic Attack (UK-TIA) aspirin trial: final results. J Neurol Neurosurg Psychiatry. 1991;54:1044 –1054.

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)*:
2010 AHA/ASA Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack, page 249.

Class I, Level of Evidence A

For patients with noncardioembolic ischemic stroke or TIA, the use of antiplatelet agents rather than oral anticoagulation is recommended to reduce the risk of recurrent stroke or another cardiovascular event.

Aspirin (50 mg/d to 325 mg/d) monotherapy (Class I, Level of Evidence A), the combination of aspirn 25 mg and extended-release dipyridamole 200 mg twice daily (Class I, Level of Evidence B), and clopidogrel 75 mg monotherapy daily (Class IIa, Level of Evidence B), are all acceptable options for initial therapy. The selection of an antiplatelet agent should be individualized on the basis of patient risk factor profiles, cost, tolerance, and other clinical characteristics.

1c.17 Clinical Practice Guideline Citation: Furie KL, Kasner SE, Adams RJ, Albers GW, et al. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2011;42:246-49.

1c.18 National Guideline Clearinghouse or other URL: http://stroke.ahajournals.org/content/42/1/227.full.pdf

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: This guideline was approved by the American Heart Association Science Advisory and Coordinating Committee on July 13, 20010. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence

from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.23 Grade Assigned to the Recommendation: Class I, Level of Evidence A

1c.24 Rationale for Using this Guideline Over Others: The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: High 1c.26 Quality: High1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0435_Evidence_MSF5.0_Data.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. The effectiveness of antithrombotic agents in reducing stroke mortality, stroke-related morbidity and recurrence rates has been studied in several large clinical trials. Studies have shown that antithrombotic therapy prescribed at discharge following acute ischemic stroke reduces stroke mortality and morbidity.

Healthcare organizations that track antithrombotic therapy prescribed at discharge for internal quality improvement purposes have seen an improvement in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The use of antithrombotic therapy after ischemic stroke is recommended for secondary stroke prevention. Earlier evidence supported substantial underutilization of this therapy, particularly among the very elderly, those admitted from skilled nursing facilities, and patients with functional dependence (Lichtman JH, 2011). An analysis of Medicare data for 31,554 non-terminally ill patients with ischemic stroke randomly selected from the Medicare Health Care Quality Improvement Program's National Stroke Project (1998 to 1999, 2000 to 2001) demonstrated that only 74.2% were discharged on an antithrombotic medication. Treatment rates decreased with age and were lowest for patients' age 85-years and older. Later data demonstrates that this previous performance gap of 25% has narrowed significantly.

In October, 2009, The Joint Commission added the stroke (STK) measure set as a new core measure option to meet performance measurement requirements for Joint Commission hospital accreditation purposes. Below is the specified level of analysis for STK-2 beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 2034 denominator cases; 1995 numerator cases; 49 hospitals; 0.98083 national aggregate rate; 0.98731 mean of hospital rates; 0.2575 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.98601 25th percentile rate/lower quartile; and, 0.94737 10th percentile rate.

CY 2010: 19,889 denominator cases; 19,622 numerator cases; 137 hospitals; 0.98658 national aggregate rate; 0.97644 mean of hospital rates; 0.05765 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99711 50th percentile rate/median rate; 0.98276 25th percentile rate/lower quartile; and, 0.93436 10th percentile rate.

CY 2011: 24,053 denominator cases; 23,812 numerator cases; 157 hospitals; 0.98998 national aggregate rate; 0.97173 mean of hospital rates; 0.10346 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99803 50th percentile rate/median rate; 0.98529 25th percentile rate/lower quartile; and, 0.94915 10th percentile rate.

CY 2012: 24,420 denominator cases; 24,217 numerator cases; 158 hospitals; 0.99169 national aggregate rate; 0.98674 mean of hospital rates; 0.04061 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.98913 25th percentile rate/lower quartile; and, 0.97059 10th percentile rate.

CY 2013: 37,661 denominator cases; 37.363 numerator cases; 262 hospitals; 0.99209 national aggregate rate; 0.9885 mean of hospital rates; 0.02736 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.98913 25th percentile rate/lower quartile; and, 0.96667 10th percentile rate.

CY 2014: 180,048 denominator cases; 178,945 numerator cases; 1299 hospitals; 0.99387 national aggregate rate; 0.98632 mean of hospital rates; 0.06957 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.99254 25th percentile rate/lower quartile; and, 0.97674 10th percentile rate.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
Not applicable.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al. 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to 75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

In the national REasons for Geographic and Racial Differences in Stroke (REGARDS) cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had

a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years of age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age : RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer, et al., 2010).

In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).

Potential disparities in secondary pharmacological prevention for recurrent stroke have been identified. A study using the uniform data system for medical rehabilitation data from nursing homes in five states found that blacks or African Americans were significantly less likely to receive antithrombotic therapy (Ottenbacher KJ, et al, 2001). In a model adjusting for age, sex, physical function, and comorbidities, blacks or African Americans proved to be 80% as likely as whites to receive antithrombotic therapy.

Based on a national cross-sectional study of nursing home residents, the use of any antithrombotic (anticoagulant or antiplatelet) therapy was lower for Hispanics (45.5%), non-Hispanic blacks (49.4%), and Asian/Pacific Islanders (38.9%) than for whites (54.3%), although higher among a subsample of Native Americans (58.0%). Aspirin use was comparable across ethnic groups, and actually higher for Native American, blacks or African Americans, and Hispanic than for whites. In contrast, oral anticoagulation (warfarin) use was significantly lower for ethnic minority groups (25.4%-31.7%), except Native Americans whose proportion was only slightly lower (36.4%) than whites (39.6%)(Christian JB, et al. 2003).

Since the last endorsement date, Schwamm and colleagues (2010) found significant and important differences in quality of care related to antithrombotics at discharge when multivariate models were constructed adjusting for patient-level characteristics only. Race/ethnicity Black versus White: unadjusted OR 0.86 [95% CI 0.83-0.90]; adjusted for patient characteristics OR 0.81 [95% CI 0.78-0.85]; adjusted for patient and hospital characteristics OR 0.88 [95% CI 0.84-0.92]. Race/ethnicity Hispanic versus White: unadjusted OR 0.85 [95% CI 0.75-0.90]; adjusted for patient characteristics OR 0.80 [95% CI 0.75-0.86]; adjusted for patient and hospital characteristics OR 0.90 [95% CI 0.82-0.97]. Total N=46,515; All n 94.98%; White 95.15%; Black 94.41%; Hispanic 94.31%.

A more recently published study (Qian F, et al, 2013) from Get With The Guidelines (GWTG) also noted racial and ethnic disparities for antithrombotics at discharge. Using patient data (n=200,900) from the American Heart Association/American Stroke Association GWTG-Stroke program from April 2003 through December 2008, the following performance measure rates were reported for discharged antithrombotic: non-Hispanic White (n=170,694) 95.3%; non-Hispanic Black (n=20,514) 94.3%; Hispanic (n=6632) 94.5%; and non-Hispanic Asian American (n=3060) 94.5%. Data from the Paul Coverdell National Acute Stroke Registry (PCNASR) were similar for antithrombotic therapy at discharge (n=58,823) White 98.4%; Other Race 98.3% (Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, 2014).

• Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, Yan T, Escarce JJ. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-8.
• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Christian JB, Lapane KL, Toppa RS. Racial disparities in receipt of secondary stroke prevention agents among nursing home residents. Stroke. 2003;34:2693-2697.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, Griffith P, Gorelick PB, Howard G, Leira EC, Morgenstern LB, Ovbiagele B, Peterson E, Rosamond W, Trimble B, Valderrama AL, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• Gillum RF, Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and

geographic variation in stroke mortality trends. Stroke. 2011;42(11):3294-6.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-75.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Jacobs BS, Birbeck G, Mullard AJ, Hickenbottom S, Kothari R, Roberts S, Reeves MJ. Quality of hospital care in African Americans and white patients with ischemic stroke and TIA. Neurology.2006;66:809-14.
• Johnston SC, Fung LH, Gillum LA, Smith WS, Brass LM, Lichtman JH, Brown AN. Utilization of intravenous tissue-type plasminogen activator for ischemic stroke at academic medical centers.: the influence of ethnicity. Stroke. 2001; 32:1061-68.
• Karve S, Balkrishnan R, Seiber E Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, Kerber KA, Skolarus LE, Meurer WJ, Burke JF, Adelman EE, Baek J, Lisabeth LD. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol. 2013;74:778–785.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Ottenbacher KJ, Smith PM, Illig SB, Fiedler RC, Gonzales V, Granger CV. Characteristics of persons rehospitalized after stroke rehabilitation. Arch Phys Med Rehabil. 2001;82:1367-1374.
• Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Schwamm LH, Syed FA, Reeves MJ, Smith EE, Saver JL, Messe S, Bhatt DL, Grau-Sepulveda MV, Peterson ED, Fonarow GC. Temporal trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With the Guidelines-Stroke hospitals. Circ Cardiovasc Quality Outcomes. 2013;6:543-549.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20

deaths in the United States remains attributable to stroke. More women than men die of stroke each year. Women accounted for almost 60% of US stroke deaths (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, et al., 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 ($P<0.05$).168 (US Burden of Disease Collaborators, 2013) . Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

Antithrombotic agents significantly reduce the incidence of a recurrent vascular event after a stroke. The effectiveness of antithrombotic agents in reducing stroke mortality, stroke-related morbidity and recurrence rates has been studied in several large clinical trials (Kernan WN, et al, 2014; Sandercock P, et al, 2014)). While the use of these agents for patients with acute ischemic stroke and transient ischemic attacks continues to be the subject of study, substantial evidence is available from completed studies. Data at this time suggest that antithrombotic therapy should be prescribed at discharge following acute ischemic stroke to reduce stroke mortality and morbidity as long as no contraindications exist.

1c.4. Citations for data demonstrating high priority provided in 1a.3

• Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A Guideline From the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1655-1711.
• Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and Thrombolytic Therapy for Ischemic Stroke. Chest Vol. 119, 2001: 300-320.
• Brott TG, Clark WM, Grotta JC, et al. Stroke the first hours. Guidelines for acute treatment. Consensus Statement. National Stroke Association. 2000.
• Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Chen ZM, Sandercock P, Pan HC, et al. Indications for early aspirin use in acute ischemic stroke: a combined analysis of 40,000 randomized patients from the Chinese acute stroke trial and the international stroke trial. On behalf of the CAST and IST collaborative groups, Stroke 2000;31:1240-1249.
• Coull BM, Williams LS, Goldstein LB, et al. Anticoagulants and Antiplatelet Agents in Acute Ischemic Stroke. Report of the Joint Stroke Guideline Development Committee of the American Academy of Neurology and the American Stroke Association (a Division of the American Heart Association) Stroke. 2002;33:1934 -1942.
• Furie KL, Kasner SE, Adams RJ, Albers GW, et al. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2011;42:241-43.
• George M, Xin T, McGruder H, Yoon P, Rosamond W, Winquist A, Hinchey J, Wall H, Pandey D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Guideline on the Use of Aspirin as Secondary Prophylaxis for Vascular Disease in Primary Care, Centre for Health Services Research University of Newcastle upon Tyne, & Centre for Health Economics of York, 1998.
• Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, Fang MC, Fisher M, Furie KL, Heck DV, Johnston SC, Kasner SE, Kittner SJ, mitchell PH, Rich MW, Richardson D, Schwamm LH, Wilson JA. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2014;45:241-2160-2236.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e-78-82, e-119-127.
• Sandercock P, Counsell C, Tseng MC, Cecconi E. Oral antiplatelet therapy for acute ischaemic stroke. The Cochrane Database of

Systematic Reviews. 2014;3:CD000029

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Appendix_A.1-635876076083056831.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Since the last endorsement date, the measure rationale for STK-02: Discharged on Antithrombotic Therapy was updated to address the use of novel oral anticoagulant drugs in stroke patients requiring antithrombotic therapy. In recent years, direct oral anticoagulant agents (DOACs) [A.K.A. novel oral anticoagulant agents (NOACs)] have been developed and approved by the U.S. Food and Drug Administration (FDA) for stroke prevention, and may be considered as an alternative to warfarin for select patients. Following FDA new drug approval of these agents, several DOACs were added to the medication table for antithrombotic medications (Appendix C, Table 8.2) used for abstraction of the measure.
All ICD-9-CM diagnosis codes and ICD-9-CM procedure codes were converted to ICD-10-CM diagnosis and ICD-10-PCS procedure codes throughout the measure specifications. The Department of Health and Human Services (HHS) mandated that all entities covered by the Health Insurance Portability and Accountability Act (HIPAA) must transition to a new set of codes for electronic health care transactions on October 1, 2015.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Ischemic stroke patients prescribed antithrombotic therapy at hospital discharge

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome

should be described in the calculation algorithm.
One data element is used to calculate the numerator:
• Antithrombotic Therapy Prescribed at Discharge – Documentation that antithrombotic therapy was prescribed at hospital discharge. Allowable values: Yes, No/UTD or unable to determine from medical record documentation.

Patients are eligible for the numerator population when the allowable value equals "yes" for the data element.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Ischemic stroke patients

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Nine data elements are used to calculate the denominator:

1. Admission Date – The month, day and year of admission to acute inpatient care.
2. Birthdate - The month, day and year the patient was born.
3. Clinical Trial - Documentation that during this hospital stay the patient was enrolled in a clinical trial in which patients with stroke were being studied. Allowable values: Yes or No/UTD.
4. Comfort Measures Only – The earliest day the physician/APN/PA documented comfort measures only after hospital arrival. Allowable values: 1 (Day 0 or 1); 2 (Day 2 or after); 3 (Timing Unclear); 4 (Not Documented/UTD).
5. Discharge Date – The month day and year the patient was discharged from acute care, left against medical advice or expired during the stay.
6. Discharge Disposition – The place or setting to which the patient was discharged on the day of hospital discharge.
7. Elective Carotid Intervention – Documentation demonstrates that the current admission is solely for the performance of an elective carotid intervention (e.g., elective carotid endarterectomy, angioplasty, carotid stenting).
Allowable values: Yes or No/UTD.
8. ICD-10-CM Principal Diagnosis Code - The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.
9. Reason For Not Prescribing Antithrombotic Therapy at Discharge – Documentation of a reason for not prescribing antithrombotic therapy at discharge.
Allowable values: Yes or No/UTD.

Population: Discharges with ICD-10-CM Principal Diagnosis Code for ischemic stroke as defined in Appendix A, Table 8.1.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
• Less than 18 years of age
• Length of Stay > 120 days
• Comfort measures only documented
• Enrolled in clinical trials related to stroke
• Admitted for elective carotid intervention
• Discharged to another hospital
• Left against medical advice
• Expired
• Discharged to home for hospice care
• Discharged to a health care facility for hospice care
• Documented reason for not prescribing antithrombotic therapy at discharge

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
• The patient age in years is equal to the Discharge Date minus the Birthdate. Patients less than 18 years are excluded.

• The Length of Stay (LOS) in days is equal to the Discharge Date minus the Admission Date. If the LOS is greater than 120 days, the patient is excluded.
• Patients with Comfort Measures Only allowable value of 1 (Day 0 or 1), 2 (Day 2 or after), and 3 (Timing unclear) are excluded.
• Patients are excluded if "Yes" is selected for Clinical Trial.
• Patients are excluded with ICD-10-PCS procedure codes for carotid intervention procedures as identified in Appendix A, Table 8.3, if medical record documentation states that the patient was admitted for the elective performance of this procedure.
• Patients with Discharge Disposition allowable value of 2 (Hospice-Home), 3 (Hospice-Health Care Facility), 4 (Acute Care Facility), 6 (Expired), or 7 (Left Against Medical Advice/AMA) are excluded.
• Patients are excluded if "Yes" is selected for Reason For Not Prescribing Antithrombotic Therapy at Discharge.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable, the measure is not stratified.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
1. Start processing. Run cases that are included in the Stroke (STK) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check ICD-10-CM Principal Diagnosis Code
a. If the ICD-10-CM Principal Diagnosis Code is not on Table 8.1, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
b. If the ICD-10-CM Principal Diagnosis Code is on Table 8.1, continue processing and proceed to Discharge Disposition.

3. Check Discharge Disposition
a. If Discharge Disposition equals 2, 3, 4, 6, 7, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
b. If Discharge Disposition equals 1, 5, 8 continue processing and proceed to Comfort Measures Only.

4. Check Comfort Measures Only

a. If Comfort Measures Only is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Comfort Measures Only equals 1, 2, or 3, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Comfort Measures Only equals 4, continue processing and proceed to Clinical Trial.

5. Check Clinical Trial
a. If Clinical Trial is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Clinical Trial equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Clinical Trial equals No, continue processing and proceed to Elective Carotid Intervention.

6. Check admitted for Elective Carotid Intervention
a. If Elective Carotid Intervention is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Elective Carotid Intervention equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Elective Carotid Intervention equals No, continue processing and proceed to Antithrombotic Therapy Prescribed at Discharge.

7. Check Antithrombotic Therapy Prescribed at Discharge
a. If Antithrombotic Therapy Prescribed at Discharge is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Antithrombotic Therapy Prescribed at Discharge equals Yes, the case will proceed to a Measure Category Assignment of E and will be in the Numerator Population. Stop processing.
c. If Antithrombotic Therapy Prescribed at Discharge equals No, continue processing and check Reason for Not Prescribing Antithrombotic Therapy at Discharge.

8. Check Reason for Not Prescribing Antithrombotic Therapy at Discharge
a. If Reason for Not Prescribing Antithrombotic Therapy at Discharge is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Reason for Not Prescribing Antithrombotic Therapy at Discharge equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Reason for Not Prescribing Antithrombotic Therapy at Discharge equals No, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter for the measure set cannot sample.

Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

Quarterly Sampling
The Quarterly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing quarterly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If "N" >/= 900, then "n" 180
If "N" 226-899, then "n" 20% of Initial Patient Population size
If "N" 45-225, then "n" 45

If “N” 6-44, No sampling; 100% Initial Patient Population required
If “N” 0-5, Submission of patient level data is not required; if submission occurs, 100% Initial Patient Population required

Monthly Sampling
The Monthly Sample Size “n”, i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size “N” for the STK Measure Set. Hospitals performing monthly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If “N >/= 300, then “n” 60
If “N 76-299, then “n” 20% of Initial Patient Population size
If “N 15-75, then “n” 15
If “N < 15, No sampling; 100% Initial Patient Population required

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Each data element in the data dictionary includes suggested data sources. The data are collected using contracted Performance Measurement Systems (vendors) that develop data collection tools based on the measure specifications. The tools are verified and tested by Joint Commission staff to confirm the accuracy and conformance of the data collection tool with the measure specifications. The vendor may not offer the measure set to hospitals until verification has been passed.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0435_MeasureTesting_MSF5.0_Data-635905391748808958.doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0435 NQF Project: Neurology Project

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability*.)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009. It is a requirement of participation in the ORYX initiative that data on all measures in the set are collected. Demographics of organizations collecting and reporting data on this measure are as follows:
170 health care organizations representing various hospital demographics:
17 For Profit; 144 Not for Profit; 9 Government
62 >=300 beds; 71 100-299 beds; 37 <100 beds
142 Urban; 28 Rural
22 Teaching; 148 Non-teaching
States represented in this data collection effort include: AL, AR, AZ, CA, CT, FL, GA, IA, ID, IL, IN, KY, LA, MA, MD, MI, MN, MO, MS, NC, NE, NJ, NM, NY, OH, PA, PR, SC, SD, TN, TX, VA, WA, WI
15 performance measurement systems are used for data transmission to The Joint Commission.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with monthly data transmission to The Joint Commission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of approximately 900 patient records identified at 30 pilot sites over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rates of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify potential measure enhancements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures.) The following is a list of the major tests conducted on the submitted data for this measure:

- Transmission of complete data
- Usage of data received (to understand if the healthcare organization provides the relevant service to treat the relevant population)
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers

- Data integrity
- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors is as follows:
- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Data element agreement rates reported to The Joint Commission for the time period of one year (4Q2010 – 3Q2011) have shown an overall agreement rate of 97.61%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-2.

Data Elements	Total 'n' numerator	Total 'n' denominator	Rate
Antithrombotic Therapy Prescribed at Discharge	442	460	96.1%
Clinical Trial	712	714	99.7%
Comfort Measures Only	709	714	99.3%
Elective Carotid Intervention	711	714	99.6%
Reason for Not Prescribing Antithrombotic Therapy at Discharge	10	11	91.0%

These agreement rates are considered to be well within acceptable levels.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* are consistent with the evidence cited in support of the measure focus (*criterion 1c)* and identify any differences from the evidence:
This measure focuses on determining the proportion of ischemic stroke patients who received a prescription for antithrombotic therapy at hospital discharge. Antithrombotic medications that qualify for inclusion in the numerator are detailed in medication Table 8.2 Antithrombotic Medications-Stroke. Medications other than those named cannot be included in the numerator. Medication tables are reviewed by a PharmD and revised twice a year. Patients less than 18 years of age that have a length of stay (LOS) more than 120 days, enrolled in a clinical trial for stroke or who were designated "comfort measures only" anytime during hospitalization are excluded. In addition, patients who were discharged to a health care facility for hospice care, home for hospice care, who expired or who left against medical advice are excluded to harmonize with other CMS/Joint Commission measures.

Operationally, there are two differences between the measure specifications and guideline recommendations. First, patients with a planned admission to the hospital for an elective carotid intervention which restores blood flow to the brain and prevents stroke, such as carotid endarterectomy or carotid stenting, are excluded from the measure. Second, patients with a contraindication to antithrombotic therapy or a documented reason why therapy is not indicated are excluded from the measure.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:
At the time this measure was originally tested measure validity was assessed via survey and focus groups of hospitals participating in the pilot test. All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant.

Since the measure has been in national use, continued validity of the measure has been determined through analysis of feedback from measure users. The Joint Commission provides a web-based application with which measure users can provide feedback regarding

appropriateness of measure specifications, request clarification of specifications, and/or provide other comments pertinent to the measure. This feedback is systematically, continually, reviewed in order to identify trends and to identify areas of the measure specifications that require clarification or revision. Additionally, Joint Commission staff continually monitors the national literature and environment in order to assess continued validity of this measure. An expert technical advisory panel (TAP) meets on a quarterly basis in order to assess continued validity of this measure. And finally, the conversion from ICD-9-CM diagnosis codes to ICD-10-CM diagnosis codes has been completed and reviewed by the Technical Advisory Panel for validity. The panel has determined that the intent of the measure has not changed as a result of the conversion. The crosswalk will also be posted in a future version of the specifications manual for public comment, and results of feedback will be reviewed and incorporated into the measure specifications where indicated.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:
Analysis of feedback obtained via our automated feedback system reveals 43 submissions regarding specifications for this measure over the past year. Predominantly, questions involve the data element Reason for Not Prescribing Antithrombotic Therapy at Discharge. Reasons for not prescribing antithrombotic therapy must be explicitly documented by the physician, advanced practice nurse, physician assistant, or pharmacists; however, medical record documentation does not always reflect the linkage of a reason with an antithrombotic drug. Questions about the acceptability of reason documentation are also common. In addition, many questions have been received regarding the acceptability of new anticoagulant agents, (i.e., dabigatran, rivaroxaban) for inclusion in the numerator population when prescribed at discharge. Both medications have been added to medication Table 8.2 Antithrombotic Medications-Stroke.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:
Measure exclusions that were not derived directly from the evidence are presented below. Please note that these are population exclusions that are necessary to ensure consistency in all measures in this 8 measure set.
These exclusions were analyzed for frequency of occurrence. An issue that is of great concern to users of this measure is that due to the presence of exceptions to the measure, attainment of a 100% measure rate is not possible. Because of the role of this measure in the current Joint Commission accreditation process this is especially troubling to measure users. This concern is the basis for a number of the non-evidence-based exclusions to these measures. The following measure exclusions that were not derived directly from the evidence are as follows:

1. Patients < 18 years old
2. Patients who have a length of stay (LOS) greater than 120 days
3. Patients with Comfort Measures Only documented
4. Patients enrolled in clinical trials
5. Patients admitted for Elective Carotid Intervention
6. Patients discharged to another hospital (acute care facility)
7. Patients who left against medical advice
8. Patients who expired
9. Patients discharged to home for hospice care
10. Patients discharged to a health care facility for hospice care
11. Patients with a documented Reason For Not Prescribing Antithrombotic Therapy at Discharge

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
N=39,812

1. Patients < 18 years old = 0%
2. Patients who have a length of stay (LOS) greater than 120 days = 0%
3. Patients with Comfort Measures Only documented = 11.37%

4. Patients enrolled in clinical trials = 0.40%
5. Patients admitted for Elective Carotid Intervention = 10.19%
6. Patients discharged to another hospital (acute care facility) = 1.16%
7. Patients who left against medical advice = 0.26%
8. Patients who expired = 3.66%
9. Patients discharged to home for hospice care = 0.61%
10. Patients discharged to a health care facility for hospice care = 0.68%
11. Patients with a documented Reason For Not Prescribing Antithrombotic Therapy at Discharge = 10.15%

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Not Applicable

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:
Not Applicable

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:
Not Applicable

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:
Not Applicable

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:
The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
STK-2 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10th Percentile
3Q2011	154	0.98351	0.07256	1	1	1	1	0.97674
2Q2011	155	0.97524	0.11930	1	1	1	1	0.96491

1Q2011	160	0.97249	0.10725	1	1	1	0.98833	0.94495
4Q2010	135	0.98048	0.06361	1	1	1	1	0.94286
3Q2010	130	0.96843	0.11476	1	1	1	0.98889	0.94291
2Q2010	122	0.97233	0.08071	1	1	1	0.97959	0.94203
1Q2010	99	0.97384	0.06745	1	1	1	0.97959	0.91667
4Q2009	49	0.98731	0.02575	1	1	1	0.98601	0.94737

It should be noted that since data collection on this measure is completely voluntary for both The Joint Commission and PCNASR, the hospitals reporting on this measure are self-selected, and therefore, presumably have a particular interest and concern for improving stroke care. For this reason, the measure results demonstrated by this group most likely significantly overstates the rate for the population of all health care organizations.

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Multiple data sources are not used for this measure.

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:
Not applicable

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:
Not applicable

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: The measure is not stratified for disparities.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:
Although some evidence exists that minorities are less likely to receive antithrombotic therapy, there are no plans to stratify the measure. The Joint Commission does not currently capture gender-specific data, or data elements for race or ethnicity because these data elements have not been shown to be reliably collectable due to the fact that no national standardized definitions exist for these data elements. Also, not all hospitals collect race and ethnicity. In the future, it may be feasible for The Joint Commission to explore how race and ethnicity and other relevant disparity data might be collected reliably in the future.

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?
(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

NEW (since 2012 endorsement): Data for Empirical Testing

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

2b2. VALIDITY TESTING
2b2.1. What level of validity testing was conducted? (*may be one or both levels*)
☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☒ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The data used to measure validity consists of one year of data (third, and fourth quarter 2014 and first and second quarter for 2015) extracted from The Joint Commission warehouse. The hospital selection was based on those hospitals that reported 12 months of data and had 30 or more denominator cases for the year.

Measure convergent and criterion validity was assessed using hospital patient level data from The Joint commission warehouse. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous face validity testing. The patient population was comprised of all patients discharged with an ICD-9-CM Principal Diagnosis Code for stroke as defined in Appendix A, Table 8.1 or Table 8.2.

Conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications. The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015.

We conducted a secondary analysis to help interpret results; correlation with other stroke (STK) process measures.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Descriptive statistics for the STK measures:
N=1,318 hospitals
n= 2, 206,379 patients records

STK-2
Median: 100%
Percentile 10%: 98%
Percentile 25%: 99%
Percentile 75%: 100%
Percentile 90%: 100%

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations							
	STK_1	STK_2	STK_3	STK_4	STK_5	STK_6	STK_10
STK_1	1.00000 1318	0.55423 <.0001 1318	0.30311 <.0001 1302	0.37009 <.0001 1144	0.52785 <.0001 1318	0.53082 <.0001 1318	0.45291 <.0001 1318
STK_2	0.55423 <.0001 1318	1.00000 1318	0.37566 <.0001 1302	0.28169 <.0001 1144	0.53355 <.0001 1318	0.47326 <.0001 1318	0.29578 <.0001 1318
STK_3	0.30311 <.0001 1302	0.37566 <.0001 1302	1.00000 1302	0.25551 <.0001 1137	0.25665 <.0001 1302	0.28198 <.0001 1302	0.15819 <.0001 1302
STK_4	0.37009 <.0001 1144	0.28169 <.0001 1144	0.25551 <.0001 1137	1.00000 1144	0.22516 <.0001 1144	0.43717 <.0001 1144	0.41076 <.0001 1144
STK_5	0.52785 <.0001 1318	0.53355 <.0001 1318	0.25665 <.0001 1302	0.22516 <.0001 1144	1.00000 1318	0.35302 <.0001 1318	0.39079 <.0001 1318
STK_6	0.53082 <.0001 1318	0.47326 <.0001 1318	0.28198 <.0001 1302	0.43717 <.0001 1144	0.35302 <.0001 1318	1.00000 1318	0.45420 <.0001 1318
STK_10	0.45291 <.0001 1318	0.29578 <.0001 1318	0.15819 <.0001 1302	0.41076 <.0001 1144	0.39079 <.0001 1318	0.45420 <.0001 1318	1.00000 1318

The correlations of the stroke measures are all positively correlated and statistically significant indicating that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)
Overall, the positive inter-correlations indicate convergent validity of the measures.
STK-02 is positively correlated with other stroke evidence-based processes of care.

2b3. EXCLUSIONS ANALYSIS .
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patient who have a Length of Stay greater than 120 days
- Patients with *Comfort Measures*
- Patient enrolled in a Clinical Trial

- Patients admitted for *Elective Carotid Intervention*
- Patients with a documented *Reason For Not Prescribing Antithrombotic Therapy at Discharge*
- Patients discharged to another hospital
- Patients who left against medical advice
- Patients who expired
- Patients discharged to home for hospice care
- Patients discharged to a health care facility for hospice care

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are detailed below.

Exclusion: *Comfort Measures Only*
Overall Occurrence n = 163,225
Overall Occurrence Percentage: 10.4%
Minimum: 0.30%
10th Percentile: 5.26%
Median: 10%
90th Percentile: 15.8%
Maximum: 35.7%

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%
10th Percentile: 0.24%
Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for *Elective Carotid Intervention*
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.5%
90th Percentile: 25.3%
Maximum: 95.2%

Exclusion: Patients with a documented *Reason For Not Prescribing Antithrombotic Therapy at Discharge*
Overall Occurrence n = 19,751
Overall Occurrence Percentage: 6.27%
Minimum: 0.27%
10th Percentile 0.92%
Median: 2.82%
90th Percentile: 16.1%
Maximum: 60.6%

Exclusion: *Discharge Disposition* - Patients discharged to another hospital
Overall Occurrence n = 449,924
Overall Occurrence Percentage: 35.7%
Minimum: 0.787%

10th Percentile: 25%
Median: 35.4%
90th Percentile: 46%
Maximum: 76.2%

Exclusion: *Discharge Disposition* - Patients who left against medical advice
Overall Occurrence n = 8,396
Overall Occurrence Percentage: 0.67%
Minimum: 0.067%
10th Percentile: 0.34%
Median: 0.86%
90th Percentile: 2.4%
Maximum: 9.67%

Exclusion: *Discharge Disposition* - Patients who expired
Overall Occurrence n = 76,168
Overall Occurrence Percentage: 6.04%
Minimum: 0.398%
10th Percentile: 1.9%
Median: 5.08%
90th Percentile: 10.2%
Maximum: 20.1%

Exclusion: *Discharge Disposition* - Patients discharged to home for hospice care
Overall Occurrence n = 658,264
Overall Occurrence Percentage: 52.2%
Minimum: 6.25%
10th Percentile: 39%
Median: 51.9%
90th Percentile: 64%
Maximum: 94.3%

Exclusion: *Discharge Disposition* - Patients discharged to a health care facility for hospice care
Overall Occurrence n = 37,804
Overall Occurrence Percentage: 3%
Minimum: 0.169%
10th Percentile: 0.87%
Median: 3.01%
90th Percentile: 6.4%
Maximum: 23%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The median frequency of exclusions range from low to moderate. The distribution of exclusions across hospitals is generally narrow indicating that the occurrence is random and likely would not bias performance

results, although the percentage of patients excluded may differ depending on whether the hospital was a stroke center.

For criterion validity, it is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a Length of Stay greater than 120 days
Rationale: Included to harmonize with other CMS/Joint Commission aligned measures.

Patients with *Comfort Measures Only* documented
Rationale: It is inappropriate to treat such patients beyond those measures necessary for comfort.

Patients enrolled in a Clinical Trial
Rationale: Antithrombotic therapy as specified by this measure may compromise a clinical trial.

Patients admitted for *Elective Carotid Intervention*
Rationale: Patients who are not admitted for treatment of an acute stroke may not require antithrombotic therapy.

Patients with a documented Reason For Not Prescribing Antithrombotic Therapy at Discharge
Rationale: It is inappropriate to treat patients who have a documented reason or contraindication to antithrombotic therapy.

Patients discharged to another hospital
Rationale: This measure is meant for patients discharged to home or health care facility other than an acute care hospital.

Patients who left against medical advice
Rationale: Hospitals do not have opportunity for provision of quality care for the non-compliant patient.

Patients who expired
Rationale: Patients who expire are not eligible to be in this measure.

Patients discharged to home for hospice care
Rationale: Antithrombotic therapy may not be warranted for the hospice patient.

Patients discharged to a healthcare facility for hospice care
Rationale: Antithrombotic therapy may not be warranted for the hospice patient.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other

If other: Data element allowable values are selected, either manually or electronically, from clinical and coded data available in medical record documentation. All medical record documentation is used in the abstraction process. Vendor data collection tools are used to import data elements needed for measure rate calculation.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract the electronic version of this measure, so continues to offer this chart-abstracted version which allows for data capture from unstructured data fields. All data elements needed to compute the STK-2 performance measure score have been retooled for capture from electronic sources. Annual updates are performed to match the eCQM specifications to the current version of the chart-abstracted specifications.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

At the present time, hospitals using this performance measure generally collect measure data via manual review of the paper medical record, the EHR or a combination of both. Collected data are submitted to The Joint Commission on a quarterly basis, by way of contracted performance measurement system vendors, as described previously. Specifications for this measure are freely available to anyone who wishes to use the measure. Feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As described above, as feedback from measure users has indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

There are no fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Quality Check® http://www.qualitycheck.org/consumer/searchQCR.aspx Hospital Compare https://www.medicare.gov/hospitalcompare/search.html Public Health/Disease Surveillance Paul Coverdell National Acute Stroke Registry http://www.cdc.gov/dhdsp/programs/stroke_registry.htm Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program http://www.jointcommission.org/ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Annual Report – Improving America's Hospitals http://www.jointcommission.org/annualreport.aspx Quality Improvement (Internal to the specific organization) Disease-Specific Care Certification for Comprehensive Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx Disease-Specific Care Certification for Primary Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

• Name of program and sponsor: Quality Check®; The Joint Commission
• Purpose: A public website that allows consumers to: search for accredited and certified organizations by city and state, by name or by zip code (up to 250 miles); find organizations by type of service provided within a geographic area; download free hospital performance measure results; and, print a list of Joint Commission certified disease-specific care programs and health care staffing firms.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

• Name of program and sponsor: Hospital Compare; Centers for Medicare & Medicaid Services
• Purpose: A public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4000+ Medicare-certified hospitals (2015)

• Name of program and sponsor: Paul Coverdell National Acute Stroke Registry; Centers for Disease Control and Prevention
• Purpose: A national registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.
• Geographic area and number and percentage of accountable entities and patients included: 11 states; 403 hospitals (CDC, 2014)

• Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
• Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 hospitals (2015)

• Name of program and sponsor: Annual Report-Improving America's Hospitals; The Joint Commission
• Purpose: The annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

• Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

• Name of program and sponsor: Disease-Specific Care Certification for Primary Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
• Geographic area and number and percentage of accountable entities and patients included Nationwide; 1079 hospitals

• Name of program and sponsor: Hospital Accreditation Program; The Joint Commission
• Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
• Geographic area and number and percentage of accountable entities and patients included Nationwide; 3300 Joint Commission-accredited hospitals (2014)

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6

years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

• Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)

The rate of antithrombotic therapy prescribed at discharge has steadily increased over the past five years based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 99.4%. A modest gap of ~3% still exists for the 10th percentile of hospitals. This trend is consistent with increases in the rate of antithrombotic therapy at discharge published by GWTG-Stroke and PCNASR. Compared to antithrombotic at discharge rates of approximately 75% at the turn of the millennium (Lichtman JH, 2011), the performance gap has significantly narrowed. Significant and important differences between races in rates of antithrombotic therapy prescribed at discharge persist (Schwamm, 2010, Qian, 3013; CDC 2014).

• Geographic area and number and percentage of accountable entities and patients included Nationwide; 1299 hospitals; 180,048 patients (The Joint Commission, 2014)

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No issues have been identified.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0068 : Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet
0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy
0438 : STK 05: Antithrombotic Therapy By End of Hospital Day Two

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy – American Academy of Neurology

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Measures 0438 Antithrombotic Therapy By End of Hospital Day 2 is the fifth (STK-5) measure in The Joint Commission stroke core measure set and also targets the ischemic stroke population; however, the timeframe for antithrombotic administration is different in this measure than STK-2. STK-5 focuses on the early management of stroke care and antithrombotic therapy administered within the first 48 hours of acute ischemic stroke onset rather than discharge. All common data elements for these measures are completely harmonized. Measure 0068 is a physician performance measure and could extend to the outpatient setting. Measure 0068 encompasses a different target population, specifically patients with ischemic vascular disease who were discharged alive for acute myocardial infarction (AMI), coronary artery bypass graft (CABG) or percutaneous coronary interventions (PCI). As previously noted, this measure evaluate physician practice as opposed to hospital processes.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not Applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Ann, Watt, awatt@jointcommission.org, 630-792-5944-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Karen, Kolbusz, kkolbusz@jointcommission.org, 630-792-5931-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel (TAP) is to provide advisory oversight in literature review, measure content and maintenance of the specifications.

Members are:

Harold P. Adams, Jr., MD
University of Iowa Health Care
Iowa City, IA

Mark J. Alberts, MD
University of Texas Southwestern
Dallas, TX

Anne W. Alexandrov, RN
Health Outcomes Institute
Fountain Hills, AZ

Nadine Allyn, RD
American Heart Association
Dallas, Texas

Mary G. George, MD
Centers for Disease Control and Prevention
Atlanta, GA

Martin Gizzi, MD
Legacy Health System
Portland, OR

Marianna Gorbaty
Mathematica Policy Research, Inc.
Washington, DC

Judith Hinchey, MD
Tuft's Medical School
Steward St. Elizabeth's Medical Center
Boston, MA

Edward C. Jauch, MD
Medical University of South Carolina
Charleston, SC 2946

Irene Katzan, MD
Cleveland Clinic
Cleveland, OH

Kimberly E. Levasseur-Franklin, PharmD
Northwestern Memorial Hospital
Chicago, IL

Kathy Morrison, RN
Penn State Hershey Medical Center
Hershey, PA

Marilyn M. Rymer, MD
University of Kansas
Kansas City, MO

Jeffrey Saver, MD
UCLA Medical Center

Los Angeles, CA

Lee H. Schwamm, MD
Harvard Medical School
Massachusetts General Hospital
Boston, MA

Deborah, Summers, RN
St. Luke's Brain and Stroke Institute
Kansas City, MO

Osama O. Zaidat, MD
St.Vincent Mercy Hospital
Toledo, OH

Richard D. Zorowitz, MD
Medstar National Rehabilitation Network
Washington, DC

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2009
Ad.3 **Month and Year of most recent revision:** 10, 2015
Ad.4 **What is your frequency for review/update of this measure?** Biannual
Ad.5 **When is the next scheduled review/update for this measure?** 07, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0436
Measure Title: STK-03: Anticoagulation Therapy for Atrial Fibrillation/Flutter
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of ischemic stroke patients with atrial fibrillation/flutter who are prescribed anticoagulation therapy at hospital discharge.

This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-4: Thrombolytic Therapy, STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-6 Discharged on Statin Medication, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs.
Developer Rationale: Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Stroke risk is significantly increased for patients with a history of or current finding of atrial fibrillation/flutter. Antiplatelet medications alone do not significantly reduce stroke risk for this patient population. Anticoagulation through the administration of warfarin, low molecular-weight heparins, or newer anticoagulant medications approved for stroke prevention is the recommended therapy.

Multiple clinical trials have demonstrated the superior therapeutic effect of warfarin compared with placebo in the prevention of thromboembolic events among patients with nonvalvular atrial fibrillation. Based on data pooled from 5 primary prevention trials of warfarin versus control, it is possible to reduce the annual stroke rate from 4.5% for the control patients to 1.4% in patients treated with dose-adjusted warfarin. In other words, 31 ischemic strokes can be prevented each year for every 1,000 patients treated (Sacco RL, et al., 2006).

Healthcare organizations that track anticoagulation therapy for internal quality improvement purposes have seen a significant increase in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e. FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

Numerator Statement: Ischemic stroke patients prescribed anticoagulation therapy at hospital discharge
Denominator Statement: Ischemic stroke patients with documented atrial fibrillation/flutter.
Denominator Exclusions:
- Less than 18 years of age
- Length of Stay > 120 days
- Comfort measures only documented
- Enrolled in clinical trials related to stroke
- Admitted for elective carotid intervention
- Discharged to another hospital
- Left against medical advice
- Expired
- Discharged to home for hospice care
- Discharged to a health care facility for hospice care

• Documented reason for not prescribing anticoagulation therapy at discharge

Measure Type: Process
Data Source: Electronic Clinical Data, Paper Medical Records
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2008 **Most Recent Endorsement Date:** Nov 01, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Summary of prior review in 2012:

- The body of evidence consistently supports that the administration of anticoagulation therapy, unless there are contraindications, is an established and effective strategy in preventing recurrent stroke in high stroke risk atrial fibrillation patients with TIA or prior stroke.
- Class I, Level of Evidence A – 2010 AHA/ASA Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack, page 243
  - For patients with ischemic stroke or TIA with paroxysmal (intermittent) or permanent AF, anticoagulation with a vitamin K antagonist (target INR, 2.5; range 2.0 to 3.0) is recommended.
  - The Committee noted that the medical evidence to support anticoagulation therapy is not controversial;
  - however, there are some questions around the evidence for the timing of anticoagulant therapy.
- The 2012 Committee noted that the medical evidence to support anticoagulation therapy is not controversial; however, there are some questions around the evidence for the timing of anticoagulant therapy.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**
☐ **The developer provided updated evidence for this measure:**

Guidance from the Evidence Algorithm:
Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency high(Box 5) → Box 5a High

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and voting on Evidence?*

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer provides the following national trend data:

	CY 2010	CY2011	CY 2012	CY 2013	CY 2014
# hospitals	136	150	149	257	1256
# cases (den)	2,952	3,566	3,685	5,635	28,0247
National aggregate rate	0.94	0.95	0.96	0.96	0.97
Mean hospital rate	0.92	0.93	0.95	0.95	0.97
50th percentile	1.00	1.00	1.00	1.00	1.00
10th and 90th percentiles	0.67 (10th) 1.00 (90th)	0.78 (10th) 1.00 (90th)	0.83 (10th) 1.00 (90th)	0.88 (10th) 1.00 (90th)	0.89 (10th) 1.00 (90th)

- Participation in this measure has grown significantly in the past five years. National hospital performance seems to be very high with little room for improvement.
- In 2012, the Committee noted that although the average performance rate for this measure was approximately 94%, there was evidence of a performance gap for minority populations.

Disparities:

- The developer does not provide disparities data from use of this measure.
- Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age."

Questions for the Committee:

- *Performance over five years has been consistently high. How much further improvement in performance is likely using this measure?*
- *Does a gap in care still exist that warrants this national performance measure?*
- *Is there evidence that a gap in care still exits for minority populations? How can this measure be used to better understand disparities in care and outcomes for certain groups?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **Note I did this before and the response was discarded by the system. I'll do again but briefly now.

**Process measure. Direct. relates to desired outcome indirectly; measures the institution of AC after stroke where Afib is present. implied connection with desired outcome of reduction of stroke without side effects such as major bleeding.

**No new studies/information since 2012

**Strong evidence - no changes. No need for discussion or vote

1b. Performance Gap

Comments: **Limited gap. has improved in hospital performance past 5 years. perhaps some subgroup gaps but data given do not document a major gap.

**High participation 97%

**Participation rate is up - modest room for improvement. No data on disparities. The gap may not warrant the measure. Or it may be that the known disparities in stroke care account for much of the remaining performance gap - in which case the continuation of the measure is warranted to help address disparities.

1c. High Priority (previously referred to as High Impact)

Comments: **NA

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): medical record abstraction (paper or electronic). This is not an eMeasure.

Specifications:

- One data element defines the numerator and 10 data elements define the denominator.
- Developer reports that "Since the last endorsement date, the measure rationale for STK-03: Anticoagulation Therapy for Atrial Fibrillation/Flutter was updated to address the use of novel oral anticoagulant drugs in stroke patients requiring anticoagulation therapy. In recent years, direct oral anticoagulant agents (DOACs) [A.K.A. novel oral anticoagulant agents (NOACs)] have been developed and approved by the U.S. Food and Drug Administration (FDA) for stroke prevention, and may be considered as an alternative to warfarin for select patients. Following FDA new drug approval of these agents, several DOACs were added to the medication table for anticoagulant medications (Appendix C, Table 8.3) used for abstraction of the measure."
- Diagnosis and procedure codes have been converted from ICD-9 to ICD-10-CM diagnosis and ICD-10-PCS procedure codes. See attached spreadsheet Appendix A Table 8.1 (ischemic stroke).
- A conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications:
    - "The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015 developer advisory panel determined that the intent of the measure was changed by the conversion to ICD-10 and a crosswalk was posted."
- Sampling monthly or quarterly is allowed; instructions for sampling are provided.
- A web-based data collection tool has been developed but is not described. Hospitals are not required to use this tool.

- The measures is specified at the hospital level of analysis.

Questions for the Committee :

- *Are changes to the measure specifications appropriate?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Inter rater-reliability of the data elements for one year (4Q2010-3Q2011) in 77 hospitals and 739 patient records showed an overall agreement rate of 98.1%.
  - Four data elements found >95% agreement: Afib/flutter (95.5%), Clinical Trial (99.7%), Comfort Measures Only (99.3%), and Elective Carotid Intervention (99.6%). No Kappa scores were presented.

- In 2012, the Committee voiced some concern that the measure submission did not specify the agents that could be used to meet the measure.
- One Committee member noted that atrial fibrillation is under-diagnosed, and therefore this measure could potentially miss many patients who should be treated. The developer clarified that the measure includes any patient for whom atrial fibrillation is documented during the hospital stay or for whom there is any documentation of past history of atrial fibrillation or flutter.
- The Committee also expressed concern about the relatively low rate of agreement (85%) for the numerator data element reliability testing. The developer explained that some of the newer anticoagulants were not on the abstractors' lists when the reliability testing was done and this likely contributed to the lower rates of agreement between the raters.

Describe any updates to testing: No new information provided

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing: see above

Results of reliability testing : see above

Guidance from the Reliability Algorithm :

Precise specifications (Box 1) →empirical testing as specified (Box 2) → reliability testing with patient-level data elements (Box 8) → assessed all critical data elements (Box 9)→ high/moderate certainty the data elements are reliable (Box 10) → Moderate (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Questions for the Committee:

- *The developer has not provided any new reliability testing. Does the Committee agree there is no need for repeat discussion and voting on Reliability?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:
- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Face validity of the data was assessed via survey and focus groups of hospitals participating in the pilot test.
- The developer reports that "All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant."
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.
- Ongoing feedback from measure users that is systematically reviewed to identify trends and/or measure specifications that require clarification or revision. Predominantly, questions regarding clarification of the data element Afib/flutter were most common. In addition, developers received questions regarding acceptability of new anticoagulant agents.

Describe any updates to validity testing – New empirical validity testing data is presented

SUMMARY OF TESTING
Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:
- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:
- Pearson Correlation Coefficients were calculated for seven stroke measures for hospitals that reported 12 months of data and had 30 or more denominator cases for the year (2014-2015).Presumably high performing hospitals would do well on all stroke performance measures.

Validity testing results:
- Analysis of 1,318 hospitals and 2,206,379 patients records generated a table of Pearson Correlation Coefficient results that shows a statistically significant (P<.0001), positive correlation of this measure with six other stroke performance measures supporting the hypothesis that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

Questions for the Committee:
- *Is the test sample adequate to generalize for widespread implementation?*

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients are excluded from the measure for the following reasons:
  - Patients less than 18 years of age
  - Patient who have a Length of Stay greater than 120 days
  - Patients with Comfort Measures
  - Patient enrolled in a Clinical Trial
  - Patients admitted for Elective Carotid Intervention
  - Patients with a documented Reason For Not Prescribing Antithrombotic Therapy at Discharge
  - Patients discharged to another hospital
  - Patients who left against medical advice
  - Patients who expired
  - Patients discharged to home for hospice care
  - Patients discharged to a health care facility for hospice care
- New data on the frequency of exclusions is presented for 2, 206,379 admissions in 1,318 hospitals:
  - Comfort Measures Only: 10.4% (range 0.30-35.7%)
  - Clinical Trial: 0.25% (range 0.08-35.5%)
  - Patients admitted for Elective Carotid Intervention: 11.8% (range 0.16-95.2%)
  - Patients with a documented Reason For Not Prescribing Antithrombotic Therapy at Discharge: 6.27% (range 0.27%-60.6%)
  - Patients discharged to another hospital: 35.7% (range 0.79-76.2%)
  - Patients who left against medical advice: 0.67% (range 0.06-9.67%)
  - Patients who expired: 6.04% (range 0.39-20.1%)
  - Patients discharged to home for hospice care: 52.2% (range 6.25-94.3%)
  - Patients discharged to a health care facility for hospice care: 3% (range 0.16-23%)

Questions for the Committee:

- *In this sample approximately half of the patients are excluded. Does the large number of exclusions for Patients discharged to home for hospice care pose a threat to validity of the measure?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Performance data is presented above under opportunity for improvement. Complete details of the data is presented in 1b.
- The developer explains their approach: "There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations."

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods: Not Applicable

2b7. Missing Data

- The developer describes the handling of missing data in the measure specifications (S.22) but does not provide information on the frequency of missing data or potential impact on measure results.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → potential threats to validity mostly assessed (Box2) →empirical validity testing (Box 3) → measure score testing (Box 6) → sound conceptualization and hypotheses (Box 7) →moderate confidence that score are a valid indicator of quality(Box 8a) → High

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **valid measure of the process.
**No concerns
**Specifications are clear and appropriate.
**Well defined.
**Changes to specifications appropriate.

2a2. Reliability Testing

Comments: **valid. correlates with other stroke hospital measures which indicate quality care.
**New testing data adequate
**Very high - face validity and new empirical validity test. Possible issues of novel anti-coagulation agents and disparity data should be examined in the future.
**reliable
**Previous reliability testing appropriate

2b2. Validity Testing

Comments: **The exclusions omit large % of at risk population but all make sense. Institution of ac in comfort care or hospice population not likely. Single concern was not capturing ac in patients transferred.
**New data exclusions are large >50% of pts excluded
would like to discuss whether this is appropriate given the purpose of the measure
**Discharged to another hospital are high - mentioned later that this really means discharged to home?????? If so, these should also be included in the numerator.
Overall high exclusion rates.
Little to identify meaningful differences.
Missing data are handled, but no identification of frequency.
**Valid. Correlates with other stroke hospital measures which indicate quality care.
**New testing data adequate

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **reliable
**Previous reliability testing approriate
**Inter-rater 98.1%. Also high on exclusions.
**The exclusions omit large % of at risk population but all make sense. Institution of ac in comfort care or hospice population not

likely. Single concern was not capturing ac in patients transferred.
**New data exclusions are large >50% of pts excluded would like to discuss whether this is appropriate given the purpose of the measure

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is medical record abstraction: not all hospitals are able to generate the data electronically so a paper-based option is available.
- Some data elements are in electronic form.
- Data collection burden is significant.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Feasible but needs chart review to determine eligibility for the measure. stroke diagnosis and AC medication and afib dx available in EHR as extractable element .

**No concerns

**In chart or EHR. Data collection and reporting may be a burden which needs to be weighed with continued contribution to quality.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- The Joint Commission Quality Check®: Public website that allows consumers to search for accredited and certified organizations.
- CMS Hospital Compare: Public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.
- CDC Paul Coverdell National Acute Stroke Registry: National registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and

recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.

- CMS Hospital Inpatient Quality Reporting Program (IQR): The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- The Joint Commission Annual Report-Improving America's Hospitals: Annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
- The Joint Commission Disease-Specific Care Certification for Comprehensive Stroke Centers: Certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
- The Joint Commission Disease-Specific Care Certification for Primary Stroke Centers: Certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results:

- The developer provided the following progress information on 1,256 hospitals nationwide and 28,027 patients:
  - The rate of anticoagulation therapy for atrial fibrillation has steadily increased over the past five years based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 97.2%. A gap of approximately 12% still exists for the 10th percentile of hospitals. This trend is consistent with increases in the rate of anticoagulation for atrial fibrillation published by GWTG-Stroke and PCNASR.
- Non-Hispanic black patients continue to be an underserved population (Schwamm, 2010, Qian, 3013; CDC 2014). Women are also less likely to receive anticoagulation therapy for atrial fibrillation (Bushnell, et al, 2014).

Unexpected findings (positive or negative) during implementation: The developer did not identify any unexpected findings during implementation.

Potential harms:

- Novel oral anticoagulant drugs (NOACs) have been added to the list of comprehensive anticoagulant medications for stroke (Appendix C, Table 8,3) as they have become approved by the FDA. Although some adverse events have been associated with the NOACs and reported in the literature, e.g., increased incidence of myocardial infarction with dabigatran, no unintended negative consequences have been reported with the addition of the warfarin-alternative agents as a result of ongoing data collection.

Feedback :

- The developer states that feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As previously noted, measure users have indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Multiple reporting organizations.

**No concerns

**Data are reported. There is still a 12% gap in the bottom 10% so room for improvement is there, could be disparity related. Need to address Novel Anti-Coagulants and potential differences in efficacy - including possible differences is varied populations.

Criterion 5: Related and Competing Measures

Related or competing measures:

- 1525 : Atrial Fibrillation and Atrial Flutter: Chronic Anticoagulation Therapy

- 0084 : Heart Failure (HF) : Warfarin Therapy Patients with Atrial Fibrillation – AMA-PCPI – no longer NQF endorsed
- 0241 : Stroke and Stroke Rehabilitation: Anticoagulant Therapy Prescribed for Atrial Fibrillation (AF) at Discharge – AAN – no longer endorsed
- 0624 : Atrial Fibrillation - Anticoagulation Therapy – ActiveHealth Management – not NQF endorsed

Harmonization :

- The target population for measure 1525 differs from measure 0436 Anticoagulation Therapy for Atrial Fibrillation/Flutter in that it includes in the denominator population all patients age 18 years and older with a diagnosis of nonvalvular atrial fibrillation or atrial flutter whose assessment of the specified thromboembolic risk factors indicate one or more high-risk factors or more than one moderate risk factor, as determined by CHADS2 risk stratification. It is not specified for ischemic stroke patients with atrial fibrillation/flutter only.

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0436 NQF Project: Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The focus of the measure is to prevent a secondary stroke attributable to a thromboembolic event resulting from atrial fibrillation following a new ischemic stroke.
Anticoagulation therapy prescribed at discharge >> secondary stroke prevention >>decreased morbidity and mortality.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
This measure is consistent with the guidelines recommended by the American Heart Association/American Stroke Association for prevention of stroke in patients with ischemic stroke or transient ischemic attack. Both persistent atrial fibrillation and paroxysmal atrial fibrillation are potent predictors of first and recurrent stroke. Data from clinical trials show that age, recent congestive heart failure, hypertension, diabetes, and prior thromboembolism have been found to identify high-risk groups for atrial thromboembolism among patients with atrial fibrillation. Left ventricular dysfunction, left atrial size, mitral annular calcification (MAC), spontaneous echo contrast, and left atrial thrombus by echocardiography have also been shown to predict increased thromboemoblic risk. Overall, patients with prior stroke carry the highest stroke risk (RR, 2.5) (Sacco RL, et al., 2006).

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: Between 1990 and 2006, a total of 24 randomized trials (RCTs) involving patients with non-valvular atrial fibrillation were published. Collectively, these RCTs represent 20,012 participants with an average follow-up of 1.6 years, and a total exposure of about 32,800 patient-years (Fuster V, et al., 2006).

A meta-analysis of six randomized trials (Hart RG, et al.1999) conducted between 1989 and 1993 showed that adjusted dose-warfarin is highly efficacious for prevention of all stroke (both ischemic and hemorrhagic), with a risk reduction of 61% (95% CI 47% to 71%) versus placebo. The duration of follow-up was generally between 1 and 2 years; the longest was, 2.2 years. Five of the six RCTs, focused on primary prevention of thromboembolism in patients with nonvavlular atrial fibrillation, (i.e., The Copenhagen AFASAK Study (AFSAK); Stroke Prevention in Atrial Fibrillation (SPAF I); The Boston Area Anticoagulation Trial for Atrial Fibrillation (BAATAF); the Canadian Atrial Fibrillation Anticoagulation (CAFA) Study; and, the Veteran Administration Stroke Prevention in Non-Rheumatic Atrial Fibrillation (VA-SPINAF). One trial, the European Atrial Fibrillation Trial (EAFT), specifically studied secondary prevention among patients who had survived nondisabling stroke or cerebral transient ischemic attack (TIA). Findings from primary prevention studies were extrapolated and applied to secondary prevention.

A Cochrane review of anticoagulants for preventing stroke in patients with nonrheumatic atrial fibrillation and a history of stroke or transient ischemic attack published in The Cochrane Library 2011, Issue 4, focused specifically on two of the above RCTs: EAFT and the VA-SPINAF. The search strategy employed the Cochrane Stroke Group Trials Register (June 9, 2003), and contacted

researchers in the field to identify any further published or non-published studies. Selection criteria involved randomized trials comparing oral anticoagulants with control (no therapy) or placebo in people with NRAF and a previous TIA or minor ischemic stroke. Control groups on aspirin did not meet search criteria. These two trials involved 485 participants. Although VA-SPINAF was a primary prevention study, a small proportion of patients had suffered a previous stroke or TIA and the results for these patients were reported separately. The review considered the following main outcomes: fatal or non-fatal (disabling or non-disabling) recurrent stroke; all major vascular events: vascular death (including fatal bleeds), recurrent stroke (both ischemic and hemorrhagic), myocardial infarction, and systemic embolism; any intracranial bleed; and, major extracranial bleed defined as severe enough to lead to hospital admission, blood transfusion, or surgery (Saxena R, Koudstaal PJ, 2011).

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting anticoagulation for stroke prevention in patients with NRAF is high. The Cochrane Review of EAFT and VA-SPINAF determined that in patients with NRAF and a recent TIA or minor ischemic stroke, oral anticoagulation therapy almost halves the odds of serious vascular event. The odds of recurrent stroke, disabling as well as non-disabling, is decreased by two-thirds. In the EAFT group of patients without contraindications to anticoagulants, the annual rate of all vascular events was 8% in patients assigned to anticoagulants (n=225) versus 17% in patients assigned to placebo (n=214). The risk of stroke was reduced from 12% to 4% per year. In other words, 90 vascular events (mainly strokes) were prevented per 1000 patients treated with anticoagulants per year. The incidences of bleeding events (major or minor, intracranial or extracranial) on anticoagulation was low (2.8% per year versus 0.7% per year in the placebo group). In the VA-SPINAF study, four patients in the placebo group (n=25) compared to two in the anticoagulant group (n=21) suffered a recurrent stroke. The number of all vascular events was 8/21 in the warfarin group versus 11/25 in the placebo group (OR 0.78, 95% CI 0.20 to 2.9). No intracranial bleeds occurred. The recurrence of stroke of 9.3% per year in the placebo group, and the odds reduction by anticoagulant therapy, were very similar to those found in the EAFT.

The results of secondary prevention studies are highly similar to those of the five primary prevention studies of NRAF patients who did not have an ischemic stroke or TIA (n=4052). The most important difference between the primary and secondary prevention studies is the much higher incidence of recurrent stroke. The secondary prevention studies observed an annual incidence of 12% in the placebo group, which is nearly three times as much as in the placebo treated groups of the primary prevention studies (4.5% per year). This makes the value of anticoagulation for secondary prevention even more impressive in absolute terms: 90 vascular events (mainly strokes) are prevented if 1000 patients are treated for one year. Compared with aspirin, oral anticoagulants significantly decrease the risk of all strokes, ischemic strokes, and cardiovascular events with only a modest increase in bleeding risk. According to Van Walraven (2002), stroke risk for patients treated with aspirin alone was 10% per year for patients with a prior history of stroke and 2.7% for those with no such history versus 4% and 1.5% respectively for patients receiving anticoagulation therapy.

The Cochrane reviewers identified several flaws in EAFT. First, anticoagulant treatment was not blinded. Second, the results of the five primary prevention trials were published while EAFT was in progress. This could have biased both EAFT Auditing Committee and the individual investigators. Hospital admission for a major extracranial bleed might have been more likely for patients on an anticoagulant. However, all members of the Audit Committee who assessed the outcomes were absolutely blinded for the assigned study treatment. Finally, the majority of recurrent vascular events in EAFT were major, which left little room for inter-observer variation.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of evidence consistently supports that the administration of anticoagulation therapy, unless there are contraindications, is an established and effective strategy in preventing recurrent stroke in high stroke risk atrial fibrillation patients with TIA or prior stroke.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
There is strong evidence that oral anticoagulants are beneficial and safe for preventing a second stroke in people with atrial fibrillation and prior stroke. The literature examined reveals that patients with NRAF and a recent TIA or minor ischemic stroke who are prescribed oral anticoagulation therapy have a 50% decrease in the odds of a second serious vascular event. The odds of recurrent stroke, disabling as well as non-disabling, is also decreased by two-thirds. Patients with atrial fibrillation have an irregular heart beat and, if not properly anticoagulated, this can cause the formation of a blood clot in the left atrium of the heart. This clot may break away and block a cerebral artery causing a stroke. Patients who have had a stroke in the presence of NRAF have a high risk of another stroke. Anticoagulant drugs, such as warfarin, make the blood thinner and prevent formation of blood clots and hence could prevent stroke. However, anticoagulant drugs may also cause bleeding in the brain and this complication could offset any benefits.

However, examination of the literature reveals that many studies do not provide information on the balance between risk and benefit of anticoagulation therapy in the early period after stroke onset in patients with atrial fibrillation. Several studies have recommended withholding anticoagulants during the first few days post stroke, especially in cases of large infarcts. In one trial involving 449 patients with ischemic stroke and atrial fibrillation, a low-molecular-weight heparin did not reduce recurrent stroke risk during the first 14 days (Berge, et al, 2000). Another large trial of similar patients (N=3169) receiving subcutaneous unfractionated heparin, the risk of recurrent ischemic stroke within the first 14 days was reduced by 50%; however, an increase in spontaneous intracerebral hemorrhage offset the benefit
(Saxena, et al., 2001). Conclusions from these studies continue to recommend oral anticoagulation treatment as secondary prevention for those patients' who have no contraindication.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: American Heart Association. During our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, and provided information that was accessible and met the requirements set out in this measure maintenance form.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.13 Grade Assigned to the Body of Evidence: Class I, Level of Evidence A

1c.14 Summary of Controversy/Contradictory Evidence: There is some uncertainty as to when anticoagulation therapy should be initiated for patients with atrial fibrillation following a stroke. As noted above, one study found an increased risk of intracerebral hemorrhage when anticoagulant therapy was initiated within the first 14 days post stroke; however, it is generally recommended that anticoagulation therapy be initiated within the first one to two weeks. Although the absolute benefits of anticoagulation are higher in secondary prevention, it remains undisputed that the risk of stroke warrants the use of warfarin in all patients with atrial fibrillation, unless contraindicated. The benefit of unfractionated and low molecular-weight heparins over warfarin, as well as target INR value, is somewhat controversial.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- Berge E, Abdelnoor M, Nakstad PH, Sandset PM. Low molecular-weight heparin versus aspirin in patients with acute ischemic stroke and atrial fibrillation: a double-blinded randomized study. HAEST Study Group. Lancet. 2000;355:1205-10.
- Fuster V, Rydén LE, Cannom DS, Crijns HJ, Curtis AB, Ellenbogen KA, Halperin JL, Le Heusey JY, Kay GN, Lowe JE, Olsson SB, Prystowsky EN, Tamargo JL, Wann S; ACC/AHA Task Force Members; ESC Committee for Practice Guidleines. Guidelines for the management of patients with atrial fibrillation-executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines; developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. Circulation. 2006;114;700-752.
- Hart RJ, Benavente O, McBride R, et al. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. Ann Intern Med. 1999;131:492-501.
- Saxena R, Koudstaal PJ for The Cochrane Collaboration. Anticoagulants for preventing stroke in patients with nonrheumatic atrial fibrillation and a history of stroke or transient ischemic attack (review). The Cochrane Library. 2011:4: 1-13.
- Saxena R, Lewis S, Berge E, Sandercock PA, Koudstaal PJ. Risk of early death and recurrent stroke effect of heparin in 3169 patients with acute ischemic stroke and atrial fibrillation in the International Stroke Trial. Stroke. 2001;32:2333-7.
- Sacco RL, Adams R, Albers G, Alberts MJ, Benavente O, Furie K, Goldstein LB, Gorelick P, Halperin J, Harbaugh R, Johnston SC, Katzan I, Kelly-Hayes M, Kenton EJ, Marks M, Schwamm LH, Tomsick T. Guidelines for Prevention of Stroke in Patients With Ischemic Stroke or Transient Ischemic Attack: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association Council on Stroke: Co-Sponsored by the Council on Cardiovascular Radiology and Intervention. Stroke. Vol. 37, 2006:577.
- Van Walraven C, Hart RG, Singer DE, Laupacis A, Connolly S, Petersen P, Koudstaal PJ, Chang Y, Hellemons B. Oral anticoagulants vs. aspirin in nonvalvular atrial fibrillation: an individual patient meta-analysis. JAMA. 2002;288:2441-8.

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)*:
2010 AHA/ASA Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack, page 243
Class I, Level of Evidence A

For patients with ischemic stroke or TIA with paroxysmal (intermittent) or permanent AF, anticoagulation with a vitamin K antagonist (target INR, 2.5; range 2.0 to 3.0) is recommended.

1c.17 Clinical Practice Guideline Citation: Furie KL, Kasner SE, Adams RJ, Albers GW, et al. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2011;42:241-43.

1c.18 National Guideline Clearinghouse or other URL: http://stroke.ahajournals.org/content/42/1/227.full.pdf

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: This guideline was approved by the American Heart Association Science Advisory and Coordinating Committee on July 13, 2010. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.23 Grade Assigned to the Recommendation: Class I, Level of Evidence A

1c.24 Rationale for Using this Guideline Over Others: The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: High 1c.26 Quality: High1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0436_Evidence_MSF5.0_Data.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Stroke risk is significantly increased for patients with a history of or current finding of atrial fibrillation/flutter. Antiplatelet medications alone do not significantly reduce stroke risk for this patient population. Anticoagulation through the administration of warfarin, low molecular-weight heparins, or newer anticoagulant medications approved for stroke prevention is the recommended therapy.

Multiple clinical trials have demonstrated the superior therapeutic effect of warfarin compared with placebo in the prevention of thromboembolic events among patients with nonvalvular atrial fibrillation. Based on data pooled from 5 primary prevention trials of warfarin versus control, it is possible to reduce the annual stroke rate from 4.5% for the control patients to 1.4% in patients treated with dose-adjusted warfarin. In other words, 31 ischemic strokes can be prevented each year for every 1,000 patients treated (Sacco RL, et al., 2006).

Healthcare organizations that track anticoagulation therapy for internal quality improvement purposes have seen a significant increase in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e. FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
In October, 2009, The Joint Commission added the stroke (STK) measure set as a new core measure option to meet performance measurement requirements for Joint Commission hospital accreditation purposes. Below is the specified level of analysis for STK-3 beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 325 denominator cases; 303 numerator cases; 43 hospitals; 0.93231 national aggregate rate; 0.90522 mean of hospital rates; 0.22409 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.875 25th percentile rate/lower quartile; and, 0.8 10th percentile rate.

CY 2010: 2952 denominator cases; 2785 numerator cases; 136 hospitals; 0.94343 national aggregate rate; 0.92168 mean of hospital rates; 0.13488 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.90767 25th percentile rate/lower quartile; and, 0.66667 10th percentile rate.

CY 2011: 3566 denominator cases; 3381 numerator cases; 150 hospitals; 0.94812 national aggregate rate; 0.92886 mean of hospital rates; 0.13698 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.90625 25th percentile rate/lower quartile; and, 0.7735 10th percentile rate.

CY 2012: 3685 denominator cases; 3530 numerator cases; 149 hospitals; 0.95794 national aggregate rate; 0.94795 mean of hospital rates; 0.11113 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.95522 25th percentile rate/lower quartile; and, 0.83333 10th percentile rate.

CY 2013: 5635 denominator cases; 5429 numerator cases; 257 hospitals; 0.96344 national aggregate rate; 0.95363 mean of hospital

rates; 0.1165 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.95349 25th percentile rate/lower quartile; and, 0.88235 10th percentile rate.

CY 2014: 28,027 denominator cases; 27,261 numerator cases; 1256 hospitals; 0.97267 national aggregate rate; 0.96598 mean of hospital rates; 0.08548 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.96429 25th percentile rate/lower quartile; and, 0.88889 10th percentile rate.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
Not Applicable

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not Applicable

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al. 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to 75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

In the national REasons for Geographic and Racial Differences in Stroke (REGARDS) cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years if age:

RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age : RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer,et al., 2010).

In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012). Although blacks and African Americans have a lower incidence of atrial fibrillation than Non-Hispanic whites (Hajat, et al. 2001), blacks and African Americans are also less likely to undergo cardiac monitoring and noninvasive cerebrovascular testing (Mitchell, et al., 2000). In REGARDS, investigators found that blacks or African Americans were less likely to be aware that they had atrial fibrillation or to be treated with warfarin.

Furthermore, minorities are less likely to receive medications for secondary prevention. One report suggests that blacks or African Americans are less likely to have thorough diagnostic evaluation after first stroke and are less likely to receive guideline-concordant stroke preventive medications, such as warfarin or other anticoagulants. In another study which used the 2005 Behavioral Risk Factor Surveillance System (BRFSS) in 11, 862 stroke survivors, little difference was found among blacks or African Americans and non-Hispanic whites in terms of secondary prevention measures. The study found that secondary prevention measures were underutilized in both racial groups.

Studies have also noted a relationship between health literacy, particularly math skills and medication compliance. A study from Estrada and colleagues (2004), found that anticoagulation control was poorer for participants with lower literacy levels. The international normalized ratio (INR) was 32% higher for participants in the lowest literacy group versus the highest (P+0.009). Other studies have found no association between literacy and the proportion of time with the INR in the therapeutic range (OR 1.0, 95% CI 0.7 to 1.4); however, no genetic factors influencing response to anticoagulation were included in the analysis (Fang MC, et al., 2006).

Since the last endorsement date, Schwamm and colleagues (2010) reported that black patients had significantly lower adjusted odds compared with white patients of receiving anticoagulation for atrial fibrillation (OR, 0.84; 95% CI, 0.75 to 0.94). Findings from Qian and associates (2013) agreed that non-Hispanic black patients were less likely to receive anticoagulation for atrial fibrillation at discharge. Using patient data (n=200,900) from the American Heart Association/American Stroke Association Get With The Guidelines (GWTG)-Stroke program from April 2003 through December 2008, Qian reported the following performance measure rates for discharged on anticoagulation for atrial fibrillation: non-Hispanic White (n=170,694) 90.0%; non-Hispanic Black (n=20,514) 88.3%; Hispanic (n=6632) 89.0%; and non-Hispanic Asian American (n=3060) 90.3%. According to data from the Paul Coverdell National Acute Stroke Registry (PCNASR) (n=9358), patients who are not white are less likely to receive anticoagulation therapy for atrial fibrillation; White 96.2%; Other Race 94.0% (Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, 2014).

White women with atrial fibrillation, as well as women of other races with atrial fibrillation, are slightly less likely to receive anticoagulation therapy than men (88% versus 89.7%; adjusted OR, 0.93; 95% CI, 0.88–0.98) (Bushnell, 2014). The attributable risk of stroke from atrial fibrillation increases with age, from 1.5% for those aged 50 to 59 years to nearly 25% for those aged =80 years. Whites carry the highest prevalence of atrial fibrillation compared with blacks, Hispanics, Asians, and other ethnic groups. The overall number of men and women with atrial fibrillation is similar, but ˜60% of atrial fibrillation patients aged >75 years are women (15.6% of men and 20.4% of women (P<0.0001).

• Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, Yan T, Escarce JJ. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-8.
• Bushnell, C, McCullough LD, Awad IA, Chireau MV, Fedder WN, Furie KL, Howard VJ, Lichtman JH, Lisabeth LD, Piña IL, Teeves MJ, Rexrode KM, Saposnik G, Singh V, Towfighi A, Vaccarino V, Walters MR. Guidelines for the Prevention of Stroke in Women A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association The American Academy of Neurology affirms the value of this guideline as an educational tool for neurologists. Stroke. 2014;45:24-25.

• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, Griffith P, Gorelick PB, Howard G, Leira EC, Morgenstern LB, Ovbiagele B, Peterson E, Rosamond W, Trimble B, Valderrama AL, on behalf of the American Heart Association Stroke Council, Council on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• Estrada CA, Martin-Hryniewicz M, Peek BT, Collins C, Byrd JC. Literacy and numeracy skills and anticoagulation control. Am J Med Sci. 2004;328:88-93.
• Fang MC, Machtinger EL, Wang F, Schillinger D. Health literacy and anticoagulation-related outcomes among patients taking warfarin. J Gen Intern Med. 2006; 21:841-46.
• Gillum RF, Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 2011;42(11):3294-6.
• Hajat C, Dundas R, Stewart JA, Lawrence E, Rudd AG, Howard R, Wolf CD. Cerebrovascular risk factors and stroke subtypes: differences between ethnic groups. Stroke. 2001;32:37-42.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-75.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Karve S, Balkrishnan R, Seiber E Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Levine DA, Kiefe CI, Howard G, Howard VJ, Williams OD, Allison JJ. Reduced medication access: a marker for vulnerability in US stroke survivors. Stroke. 2007;38:1557-64.
• Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Mitchell JB, Ballard DJ, Atchar DB, Whisant JP, Samsa GP. Racial variation in treatment for transient ischemic attacks: impact of participation by neurologists. Heatlh Serv Res. 2000;34:1413-1428.
• Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, Kerber KA, Skolarus LE, Meurer WJ, Burke JF, Adelman EE, Baek J, Lisabeth LD. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol. 2013;74:778–785.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, Matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Ross JS, Halm EA, Bravata DM. Use of stroke secondary prevention services: are there disparities in care? Stroke. 2009;40:1811-19.
• Schwamm LH, Syed FA, Reeves MJ, Smith EE, Saver JL, Messe S, Bhatt DL, Grau-Sepulveda MV, Peterson ED, Fonarow GC. Temporal trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With the Guidelines-Stroke hospitals. Circ Cardiovasc Quality Outcomes. 2013;6:543-549
• Tuhrim S, Cooperman A, Rojas M, Brust JC, Koppel B, Martin K, Chassin M. The association of race and sex with the underuse of stroke prevention measures. J Cerebrovascular Dis. 2008;17:226-234.

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare
Affects large numbers, A leading cause of morbidity/mortality
1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.
Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States is still attributable to stroke. More women than men die of stroke each year. Women accounted for almost 60% of US stroke deaths in 2008 (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 (P<0.05), (US Burden of Disease Collaborators, 2013). Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

Approximately 20% of ischemic strokes result from a cerebral embolism secondary to a cardiac arrhythmia or disorder. Atrial fibrillation (AF) is the most common sustained cardiac rhythm disturbance (CDC, 2010). Paroxysmal, persistent, and permanent atrial fibrillation are strong predictors of first and recurrent stroke, increasing ischemic stroke risk four to five-fold. It is estimated that over 2.3 million Americans have atrial fibrillation, and the incidence becomes more prevalent with age. AF accounts for ˜ 1.5% of stroke in individuals 50 to 59 years of age to nearly 25% in those aged > 80 years (Bushnell C, et al., 2014).

Patients who have suffered an ischemic stroke who have a high-risk source of cardiogenic embolism should generally be treated with anticoagulant drugs to prevent reoccurrence. For most patients with ischemic stroke and atrial fibrillation, it is reasonable to initiate anticoagulation therapy within 14 days of stroke onset (Kernan WN, eta l, 2014). Warfarin, dabigatran, and apixaban are all indicated for the prevention of recurrent stroke in patients with nonvalvular atrial fibrillation, whether paroxysmal or permanent. Rivaroxaban is a reasonable alternative (Kernan WN, 2014). Ischemic stroke rates per 1000 patient-years declined in AF patients taking anticoagulants, from 46.7% in 1992 to 19.1% in 2002 (AHA 2012). According to the Framingham Study (1996), AF is also an independent risk factor for ischemic stroke severity, recurrence, and mortality (Lin HJ, et al., 1996). In a study from Penado and associates (2003), people who had AF and were not treated with anticoagulants had a 2.1-fold increase in risk for recurrent stroke and a 2.4-fold increase in risk for recurrent severe stroke.

In addition to the costs attributed to stroke, the treatment of atrial fibrillation alone represents a significant health care burden. The estimated cost of treatment of atrial fibrillation in 2005 was $6.65 billion per year, including the costs of hospitalization, inpatient and outpatient physician care, and medications (Roger VL, et al.,2012).

1c.4. Citations for data demonstrating high priority provided in 1a.3
• Bushnell, C, McCullough LD, Awad IA, Chireau MV, Fedder WN, Furie KL, Howard VJ, Lichtman JH, Lisabeth LD, Piña IL, Teeves MJ, Rexrode KM, Saposnik G, Singh V, Towfighi A, Vaccarino V, Walters MR, Guidelines for the Prevention of Stroke in Women A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association The American Academy of Neurology affirms the value of this guideline as an educational tool for neurologists. Stroke. 2014;45:24-25.
• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Centers for Disease Control and Prevention, Division for Heart Disease and Stroke Prevention, National Center for Chronic Disease Prevention and Health Promotion. Fact sheets and at-a-glance reports: atrial fibrillation fact sheet. 2010.
• George M, Xin T, McGruder H, Yoon P, Rosamond W, Winquist A, Hinchey J, Wall H, Pandey D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, Fang MC, Fisher M, Furie KL, Heck DV,

Johnston SC, Kasner SE, Kittner SJ, mitchell PH, Rich MW, Richardson D, Schwamm LH, Wilson JA. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2014;45:241-2160-2236.
• Lin HJ, Wolf PA, Kelly-Hayes M, Beiser AS, Kase CS, Benjamin EJ, D'Agostino RB. Stroke severity in atrial fibrillation: The Framingham Study. Stroke. 1996;27:1760-64.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Penado S, Cano M, Acha O, Hernández JL,Riancho JA. Atrial fibrillation as a risk factor for stroke recurrence. Am J Med. 2003;114:206-210.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e-78-82, e-119-127.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not Applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Appendix_A.1-635882183961489008.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Since the last endorsement date, the measure rationale for STK-03: Anticoagulation Therapy for Atrial Fibrillation/Flutter was

updated to address the use of novel oral anticoagulant drugs in stroke patients requiring anticoagulation therapy. In recent years, direct oral anticoagulant agents (DOACs) [A.K.A. novel oral anticoagulant agents (NOACs)] have been developed and approved by the U.S. Food and Drug Administration (FDA) for stroke prevention, and may be considered as an alternative to warfarin for select patients. Following FDA new drug approval of these agents, several DOACs were added to the medication table for anticoagulant medications (Appendix C, Table 8.3) used for abstraction of the measure. Abstraction guidelines for the data element Atrial Fibrillation/Flutter were also modified to address questions received from data abstractors regarding conflicting documentation, questionable episodes of atrial fibrillation, and patients discharged from the hospital with cardiac monitoring for screening purposes only.

All ICD-9-CM diagnosis codes and ICD-9-CM procedure codes were converted to ICD-10-CM diagnosis and ICD-10-PCS procedure codes throughout the measure specifications. The Department of Health and Human Services (HHS) mandated that all entities covered by the Health Insurance Portability and Accountability Act (HIPAA) must transition to a new set of codes for electronic health care transactions on October 1, 2015

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Ischemic stroke patients prescribed anticoagulation therapy at hospital discharge

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
One data element is used to calculate the numerator:

- Anticoagulation Therapy Prescribed at Discharge – Documentation that anticoagulation therapy was prescribed at hospital discharge. Allowable values: Yes, No/UTD or unable to determine from medical record documentation.

Patients are eligible for the numerator population when the allowable value equals "yes" for the data element.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Ischemic stroke patients with documented atrial fibrillation/flutter.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Ten data elements are used to calculate the denominator:

1. Admission Date – The month, day and year of admission to acute inpatient care.
2. Atrial Fibrillation/Flutter – Documentation that the patient has a history of any atrial fibrillation (e.g., remote, persistent, or paroxysmal) or atrial flutter in the past OR current atrial fibrillation or flutter on EKG.
Allowable values: Yes or No/UTD.
3. Birthdate - The month, day and year the patient was born.
4. Clinical Trial - Documentation that during this hospital stay the patient was enrolled in a clinical trial in which patients with stroke were being studied. Allowable values: Yes or No/UTD.
5. Comfort Measures Only – The earliest day the physician/APN/PA documented comfort measures only after hospital arrival.
Allowable values: 1 (Day 0 or 1); 2 (Day 2 or after); 3 (Timing Unclear); 4 (Not Documented/UTD).
6. Discharge Date – The month, day and year the patient was discharged from acute care, left against medical advice or expired during the stay.

7. Discharge Disposition – The place or setting to which the patient was discharged on the day of hospital discharge.
8. Elective Carotid Intervention – Documentation demonstrates that the current admission is solely for the performance of an elective carotid intervention (e.g., elective carotid endarterectomy, angioplasty, carotid stenting).
Allowable values: Yes or No/UTD.
9. ICD-10-CM Principal Diagnosis Code - The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.
10. Reason For Not Prescribing Anticoagulation Therapy at Discharge – Documentation of a reason for not prescribing anticoagulation therapy at discharge.
Allowable values: Yes or No/UTD.

Population: Discharges with ICD-10-CM Principal Diagnosis Code for ischemic stroke as defined in Appendix A, Table 8.1, and patients with documented Atrial Fibrillation/Flutter.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
- Less than 18 years of age
- Length of Stay > 120 days
- Comfort measures only documented
- Enrolled in clinical trials related to stroke
- Admitted for elective carotid intervention
- Discharged to another hospital
- Left against medical advice
- Expired
- Discharged to home for hospice care
- Discharged to a health care facility for hospice care
- Documented reason for not prescribing anticoagulation therapy at discharge

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
- The patient age in years is equal to the Discharge Date minus the Birthdate. Patients less than 18 years are excluded.
- The Length of Stay (LOS) in days is equal to the Discharge Date minus the Admission Date. If the LOS is greater than 120 days, the patient is excluded.
- Patients with Comfort Measures Only allowable value of 1 (Day 0 or 1), 2 (Day 2 or after), and 3 (Timing unclear) are excluded.
- Patients are excluded if "Yes" is selected for Clinical Trial.
- Patients are excluded with the following ICD-10-PCS procedure codes for carotid intervention procedures as identified in Appendix A, Table 8.3, if medical record documentation states that the patient was admitted for the elective performance of this procedure.
- Patients with Discharge Disposition allowable value of 2 (Hospice-Home), 3 (Hospice-Health Care Facility), 4 (Acute Care Facility), 6 (Expired), or 7 (Left Against Medical Advice/AMA) are excluded.
- Patients are excluded if "Yes" is selected for Reason For Not Prescribing Anticoagulation Therapy.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable, the measure is not stratified.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*

Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
Not Applicable

S.16. Type of score:
Rate/proportion
If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. Calculation Algorithm/Measure Logic *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
1. Start processing. Run cases that are included in the Stroke (STK) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check ICD-10-CM Principal Diagnosis Code
a. If the ICD-10-CM Principal Diagnosis Code is not on Table 8.1, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
b. If the ICD-10-CM Principal Diagnosis Code is on Table 8.1, continue processing and proceed to Discharge Disposition.

3. Check Discharge Disposition
a. If Discharge Disposition equals 2, 3, 4, 6, 7, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
b. If Discharge Disposition equals 1, 5, 8, continue processing and proceed to Comfort Measures Only.

4. Check Comfort Measures Only
a. If Comfort Measures Only is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Comfort Measures Only equals 1, 2, or 3, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Comfort Measures Only equals 4, continue processing and proceed to Clinical Trial.

5. Check Clinical Trial
a. If Clinical Trial is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Clinical Trial equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Clinical Trial equals No, continue processing and proceed to Elective Carotid Intervention.

6. Check admitted for Elective Carotid Intervention
a. If Elective Carotid Intervention is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Elective Carotid Intervention equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If Elective Carotid Intervention equals No, continue processing and proceed to Atrial Fibrillation/Flutter.

7. Check Atrial Fibrillation/Flutter.
a. If Atrial Fibrillation/Flutter is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Atrial Fibrillation/Flutter equals No, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If Atrial Fibrillation/Flutter equals Yes, continue processing and check Anticoagulation Therapy Prescribed at Discharge.

8. Check Anticoagulation Therapy Prescribed at Discharge.
a. If Anticoagulation Therapy Prescribed at Discharge is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Anticoagulation Therapy Prescribed at Discharge equals Yes, the case will proceed to a Measure Category Assignment of E and will be in the Numerator Population. Stop processing.
c. If Anticoagulation Therapy Prescribed at Discharge equals No, continue processing and check Reason for Not Prescribing Anticoagulation Therapy at Discharge.

9. Check Reason for Not Prescribing Anticoagulation Therapy at Discharge.
a. If Reason for Not Prescribing Anticoagulation Therapy at Discharge is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Reason for Not Prescribing Anticoagulation Therapy at Discharge equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If Reason for Not Prescribing Anticoagulation Therapy at Discharge equals No, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter for the measure set cannot sample.

Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

Quarterly Sampling
Quarterly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing quarterly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:

If "N" >/= 900, then "n" 180
If "N" 226-899, then "n" 20% of Initial Patient Population size
If "N" 45-225, then "n" 45
If "N" 6-44, No sampling,then 100% Initial Patient Population required
If "N" 0-5, Submission of patient level data is not required, if submission occurs, 100% Initial Patient Population required

Monthly Sampling
The Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size for "N" for the STK Measure Set. Hospitals performing monthly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:

If "N" >/= 300, then "n" 60
If "N" 76-299, then "n" 20% of Initial Patient Population size
If "N" 15-75, then "n" 15
If "N" < 15, No sampling;100% Initial Patient Population required

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Each data element in the data dictionary includes suggested data sources. The data are collected using contracted Performance Measurement Systems (vendors) that develop data collection tools based on the measure specifications. The tools are verified and tested by Joint Commission staff to confirm the accuracy and conformance of the data collection tool with the measure specifications. The vendor may not offer the measure set to hospitals until verification has been passed.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not Applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0436_MeasureTesting_MSF5.0_Data-635905394179647704.doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0436 NQF Project: Neurology Project

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability*.)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on this measure are as follows:

170 health care organizations representing various hospital demographics:

17 For Profit; 144 Not for Profit; 9 Government

62 >=300 beds; 71 100-299 beds; 37 <100 beds

142 Urban; 28 Rural

22 Teaching; 148 Non-teaching

States represented in this data collection effort include: AL, AR, AZ, CA, CT, FL, GA, IA, ID, IL, IN, KY, LA, MA, MD, MI, MN, MO, MS, NC, NE, NJ, NM, NY, OH, PA, PR, SC, SD, TN, TX, VA, WA, WI

15 performance measurement systems are used for data transmission to The Joint Commission.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:

At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with monthly data transmission to The Joint Commission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of approximately 900 patient records identified at 30 pilot sites over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rates of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify potential measure enhancements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of

data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures.) The following is a list of the major tests conducted on the submitted data for this measure:

- Transmission of complete data
- Usage of data received (to understand if the healthcare organization provides the relevant service to treat the relevant population)
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity
- Data corrections

Data Element Agreement Rate:

Inter-rater reliability testing methodology utilized by contracted performance measure system vendors is as follows:

- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:

Data element agreement rates reported to The Joint Commission for the time period of one year (4Q2010 to 3Q2011) have shown an overall agreement rate of 98.1%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-3.

Data Elements	total 'n' numerator	total 'n' denominator	rate
Anticoagulation Therapy Prescribed at Discharge	127	150	84.7%
Atrial Fibrillation/Flutter	446	467	95.5%
Clinical Trial	712	714	99.7%
Comfort Measures Only	709	714	99.3%
Elective Carotid Intervention	711	714	99.6%
Reason for Not Prescribing Anticoagulation			

Therapy at Discharge	38	46	82.6%

These agreement rates are considered to be well within acceptable levels.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* are consistent with the evidence cited in support of the measure focus (*criterion 1c)* and identify any differences from the evidence:

This measure focuses on determining the proportion of ischemic stroke patients with atrial fibrillation/flutter who are prescribed anticoagulation therapy at hospital discharge. These specifications are consistent with the AHA guideline recommendations for long-term anticoagulation therapy rather than antiplatelet therapy for atrial fibrillation patients with a recent ischemic stroke.

Anticoagulant medications that qualify for inclusion in the numerator are detailed in medication Table 8.3 Anticoagulant Medications for Stroke. Patients prescribed medications other than those named cannot be included in the numerator. Medication tables are reviewed by a PharmD member of the technical advisory panel (TAP) and updated twice year. Patients less than 18 years of age that have a length of stay (LOS) more than 120 days, enrolled in a clinical trial for stroke or who were designated "comfort measures only" anytime during hospitalization are excluded. In addition, patients who were discharged to a health care facility for hospice care, home for hospice care, who expired or who left against medical advice are excluded to harmonize with other CMS/Joint Commission measures.

Operationally, there are two differences between the measure specifications and guideline recommendations. First, patients with a planned admission to the hospital for an elective carotid intervention which restores blood flow to the brain and prevents stroke, such as carotid endarterectomy or carotid stenting, are excluded from the measure. Second, patients with a contraindication to anticoagulation therapy or a documented reason why therapy is not indicated are excluded from the measure.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:

At the time this measure was originally tested measure validity was assessed via survey and focus groups of hospitals participating in the pilot test. All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant.

Since the measure has been in national use, continued validity of the measure has been determined through analysis of feedback from measure users. The Joint Commission provides a web-based application with which measure users can provide feedback regarding appropriateness of measure specifications, request clarification of specifications, and/or provide other comments pertinent to the measure. This feedback is systematically, continually, reviewed in order to identify trends

and to identify areas of the measure specifications that require clarification or revision. Additionally, Joint Commission staff continually monitors the national literature and environment in order to assess continued validity of this measure. An expert technical advisory panel (TAP) meets on a quarterly basis in order to assess continued validity of this measure. And finally, the conversion from ICD-9-CM diagnosis codes to ICD-10-CM diagnosis codes has been completed and reviewed by the Technical Advisory Panel for validity. The panel has determined that the intent of the measure has not changed as a result of the conversion. The crosswalk will also be posted in a future version of the specifications manual for public comment, and results of feedback will be reviewed and incorporated into the measure specifications where indicated.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:

Analysis of feedback obtained via our automated feedback system reveals 63 submissions regarding specifications for this measure over the past year. Questions regarding clarification of the data element Atrial Fibrillation/Flutter are most common. Abstractors are instructed to select 'YES' if there is any documentation of a history of or current finding of atrial fibrillation/flutter in the medical record. Despite this abstraction guideline, questions regarding the frequency of atrial fibrillation/flutter episodes, the duration of the episode, confirmed versus non-confirmed EKG findings, conflicting documentation, and past history of ablation procedure or permanent pacemaker for atrial fibrillation are frequent themes. In addition, many questions have been received regarding the acceptability of new anticoagulant agents, (i.e., dabigatran, rivaroxaban) for inclusion in the numerator population when prescribed at discharge. Both medications have been added to medication Table 8.3 Anticoagulant Medications-Stroke.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:

Measure exclusions that were not derived directly from the evidence are presented below. Please note that these are population exclusions that are necessary to ensure consistency in all measures in this 8 measure set.

These exclusions were analyzed for frequency of occurrence. An issue that is of great concern to users of this measure is that due to the presence of exceptions to the measure, attainment of a 100% measure rate is not possible. Because of the role of this measure in the current Joint Commission accreditation process this is especially troubling to measure users. This concern is the basis for a number of the non-evidence-based exclusions to these measures. The following measure exclusions that were not derived directly from the evidence are as follows:

1. Patients < 18 years old
2. Patients who have a length of stay (LOS) greater than 120 days
3. Patients with Comfort Measures Only documented
4. Patients enrolled in clinical trials
5. Patients admitted for Elective Carotid Intervention

6. Patients discharged to another hospital (acute care facility)
7. Patients who left against medical advice
8. Patients who expired
9. Patients discharged to home for hospice care
10. Patients discharged to a health care facility for hospice care
11. Patients with a documented Reason For Not Prescribing Anticoagulation Therapy at Discharge

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*

N=39,812

1. Patients < 18 years old = 0%
2. Patients who have a length of stay (LOS) greater than 120 days = 0%
3. Patients with Comfort Measures Only documented = 11.11%
4. Patients enrolled in clinical trials =0.39%
5. Patients admitted for Elective Carotid Intervention = 9.96%
6. Patients discharged to another hospital (acute care facility) = 1.13%
7. Patients who left against medical advice = 0.25%
8. Patients who expired = 3.58%
9. Patients discharged to home for hospice care = 0.60%
10. Patients discharged to a health care facility for hospice care = 1.52%
11. Patients with a documented Reason For Not Prescribing Anticoagulation Therapy at Discharge = 10.84%

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

Not Applicable

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:

Not Applicable

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:

Not Applicable

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: Not Applicable

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:

The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of an HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:

STK-3 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10th Percentile
3Q2011	131	0.94408	0.16256	1	1	1	1	0.83333
2Q2011	131	0.94408	0.16256	1	1	1	1	0.83333
1Q2011	147	0.92518	0.17565	1	1	1	0.94737	0.66667
4Q2010	126	0.92909	0.19535	1	1	1	1	0.8
3Q2010	119	0.94605	0.15078	1	1	1	1	0.75
2Q2010	113	0.93581	0.16405	1	1	1	0.93333	0.81818
1Q2010	90	0.91729	0.15775	1	1	1	0.88889	0.66667
4Q2009	43	0.90522	0.22409	1	1	1	0.875	0.8

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data;*

if a sample, characteristics of the entities included):

Multiple data sources are not used for this measure.

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:

Not applicable

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:

Not applicable

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: The measure is not stratified for disparities.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:

Although some evidence exists that minorities are less likely to receive secondary prevention measures such as anticoagulation therapy for atrial fibrillation/flutter, there are no plans to stratify the measure. The Joint Commission does not currently capture gender-specific data, or data elements for race or ethnicity because these data elements have not been shown to be reliably collectable due to the fact that no national standardized definitions exist for these data elements. Also, not all hospitals collect race and ethnicity. In the future, it may be feasible for The Joint Commission to explore how race and ethnicity and other relevant disparity data, might be collected reliably in the future.

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?

(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐

Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

NEW (since 2012 endorsement): Data for Empirical Testing

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The data used to measure validity consists of one year data (third, and fourth quarter 2014 and first and second quarter for 2015) extracted from The Joint Commission warehouse. The hospitals selection was based to those hospitals that reported 12 months of data and had 30 or more denominator cases for the year.

Measure convergent and criterion validity was assessed using hospitals patient level data from The Joint commission warehouse. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous face validity testing. The patient population was comprised of all patients discharged with an ICD-9-CM Principal Diagnosis Code for stroke as defined in Appendix A, Table 8.1 or Table 8.2.

Conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications. The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015.

We conducted a secondary analysis to help interpret results; correlation with other stroke process measures.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Descriptive statistics for the STK measures:

N=1,318 hospitals

n= 2, 206,379 patients records

STK-3

Median: 100%

Percentile 10%: 90%

Percentile 25%: 97%

Percentile 75%: 100%

Percentile 90%: 100%

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	STK_1	STK_2	STK_3	STK_4	STK_5	STK_6	STK_10
STK_1	1.00000 1318	0.55423 <.0001 1318	0.30311 <.0001 1302	0.37009 <.0001 1144	0.52785 <.0001 1318	0.53082 <.0001 1318	0.45291 <.0001 1318
STK_2	0.55423 <.0001 1318	1.00000 1318	0.37566 <.0001 1302	0.28169 <.0001 1144	0.53355 <.0001 1318	0.47326 <.0001 1318	0.29578 <.0001 1318
STK_3	0.30311 <.0001 1302	0.37566 <.0001 1302	1.00000 1302	0.25551 <.0001 1137	0.25665 <.0001 1302	0.28198 <.0001 1302	0.15819 <.0001 1302
STK_4	0.37009 <.0001 1144	0.28169 <.0001 1144	0.25551 <.0001 1137	1.00000 1144	0.22516 <.0001 1144	0.43717 <.0001 1144	0.41076 <.0001 1144
STK_5	0.52785 <.0001 1318	0.53355 <.0001 1318	0.25665 <.0001 1302	0.22516 <.0001 1144	1.00000 1318	0.35302 <.0001 1318	0.39079 <.0001 1318
STK_6	0.53082 <.0001 1318	0.47326 <.0001 1318	0.28198 <.0001 1302	0.43717 <.0001 1144	0.35302 <.0001 1318	1.00000 1318	0.45420 <.0001 1318
STK_10	0.45291 <.0001 1318	0.29578 <.0001 1318	0.15819 <.0001 1302	0.41076 <.0001 1144	0.39079 <.0001 1318	0.45420 <.0001 1318	1.00000 1318

The correlations of the stroke measures are all positively correlated and statistically significant indicating that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)
Overall, the positive inter-correlations indicates convergent validity of the measures. STK-03 is positively correlated with other stroke evidence-based processes of care.

2b3. EXCLUSIONS ANALYSIS .

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patient who have a Length of Stay greater than 120 days
- Patients with *Comfort Measures*
- Patient enrolled in a Clinical Trial
- Patients admitted for *Elective Carotid Intervention*
- Patients with a documented *Reason For Not Prescribing Anticoagulation Therapy at Discharge*
- Patients discharged to another hospital
- Patients who left against medical advice
- Patients who expired
- Patients discharged to home for hospice care
- Patients discharged to a health care facility for hospice care

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the

2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are given below.

Exclusion: *Comfort Measures Only*
Overall Occurrence n = 163,225

Overall Occurrence Percentage 10.4%

Minimum 0.30%

10th Percentile: 5.26%

Median: 10%

90th Percentile: 15.8%

Maximum: 35.7%

Exclusion: Clinical Trial

Overall Occurrence n = 5,446

Overall Occurrence Percentage: 0.25%

Minimum: 0.08%

10th Percentile: 0.24%

Median: 0.52%

90th Percentile: 1.90%

Maximum: 10%

Exclusion: Patients admitted for *Elective Carotid Intervention*

Overall Occurrence n = 259,833

Overall Occurrence Percentage: 11.8%

Minimum: 0.16%

10th Percentile: 2.28%

Median: 11.5%

90th Percentile: 25.3%

Maximum: 95.2%

Exclusion: Patients with a documented *Reason For Not Prescribing Anticoagulation Therapy at Discharge*

Overall Occurrence n = 25,249

Overall Occurrence Percentage 8.01%

Minimum: 0.25%

10th Percentile: 1.96%

Median: 5.22%

90th Percentile: 16%

Maximum 84.46%

Exclusion: *Discharge Disposition* - Patients discharged to another hospital

Overall Occurrence n = 449,924

Overall Occurrence Percentage 35.7%

Minimum: 0.787%

10th Percentile: 25%

Median: 35.4%

90th Percentile: 46%

Maximum: 76.2%

Exclusion: *Discharge Disposition* - Patients who left against medical advice
Overall Occurrence n = 8,396
Overall Occurrence Percentage 0.67%
Minimum: 0.067%
10th Percentile: 0.34%
Median: 0.86%
90th Percentile: 2.4%
Maximum: 9.67%

Exclusion: *Discharge Disposition* - Patients who expired
Overall Occurrence n = 76,168
Overall Occurrence Percentage 6.04%
Minimum: 0.398%
10th Percentile: 1.9%
Median: 5.08%
90th Percentile: 10.2%
Maximum: 20.1%

Exclusion: *Discharge Disposition*- Patients discharged to home for hospice care
Overall Occurrence n = 658,264
Overall Occurrence Percentage 52.2%
Minimum: 6.25%
10th Percentile: 39%

Median: 51.9%
90th Percentile: 64%
Maximum 94.3%

Exclusion: *Discharge Disposition* - Patients discharged to a health care facility for hospice care
Overall Occurrence n = 37,804
Overall Occurrence Percentage 3%
Minimum: 0.169%
10th Percentile: 0.87 %
Median: 3.01%

90th Percentile: 6.4%

Maximum: 23%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The median frequency of exclusions range from low to moderate. The distribution of exclusions across hospitals is generally narrow indicating that the occurrence is random and likely would not bias performance results, although the percentage of patients excluded may differ depending on whether the hospital was a stroke center.

For criterion validity, it is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age

Rationale: The literature does not support use in the pediatric patient.

Patients who have a Length of Stay greater than 120 days

Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Patients with *Comfort Measures Only* documented

Rationale: It is inappropriate to treat such patients beyond those measures necessary for comfort.

Patients enrolled in a Clinical Trial

Rationale: Anticoagulation therapy as specified by this measure may compromise a clinical trial.

Patients admitted for *Elective Carotid Intervention*
Rationale: Patients who are not admitted for treatment of an acute stroke may not require anticoagulation therapy

Patients with a documented *Reason For Not Prescribing Anticoagulation Therapy at Discharge*

Rationale: It is inappropriate to treat patients who have a documented reason or contraindication to anticoagulation therapy

Patients discharged to another hospital

Rationale: This measure is meant for patients discharged to home.

Patients who left against medical advice

Rationale: Hospitals do not have opportunity for provision of quality care for the non-compliant patient.

Patients who expired

Rationale: Patients who expire are not eligible to be in this measure

Patients discharged to home for hospice care

Rationale: Anticoagulation therapy may not be warranted for the hospice patient

Patients discharged to a healthcare facility for hospice care

Rationale: Anticoagulation therapy may not be warranted for the hospice patient.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other

If other: Data element allowable values are selected, either manually or electronically, from clinical and coded data available in medical record documentation. All medical record documentation is used in the abstraction process. Vendor data collection tools are used to import data elements needed for measure rate calculation.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract the electronic version of this measure, so continues to offer this chart-abstracted version which allows for data capture from unstructured data fields. All data elements needed to compute the STK-3 performance measure score have been retooled for capture from electronic sources. Annual updates are performed to match the eCQM specifications to the current version of the chart-abstracted specifications.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

At the present time, hospitals using this performance measure generally collect measure data via manual review of the paper medical record, the EHR or a combination of both. Collected data are submitted to The Joint Commission on a quarterly basis, by way of contracted performance measurement system vendors, as described previously. Specifications for this measure are freely available to anyone who wishes to use the measure. Feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As described above, as feedback from measure users has indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

There are no fees or licensing requirements to use The Joint Commission performance measures, all of which are in the public

domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting http://www.qualitycheck.org/consumer/searchQCR.aspx Quality Check® Hospital Compare https://www.medicare.gov/hospitalcompare/search.html Public Health/Disease Surveillance Paul Coverdell National Acute Stroke Registry http://www.cdc.gov/dhdsp/programs/stroke_registry.htm Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program http://www.jointcommission.org/ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Annual Report – Improving America's Hospitals http://www.jointcommission.org/annualreport.aspx Quality Improvement (Internal to the specific organization) Disease-Specific Care Certification for Comprehensive Stroke Centers Disease-Specific Care Certification for Primary Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

- Name of program and sponsor: Quality Check®; The Joint Commission
- Purpose: A public website that allows consumers to: search for accredited and certified organizations by city and state, by name or by zip code (up to 250 miles); find organizations by type of service provided within a geographic area; download free hospital performance measure results; and, print a list of Joint Commission certified disease-specific care programs and health care staffing firms.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint

Commission-accredited hospitals (2014)

- Name of program and sponsor: Hospital Compare; Centers for Medicare & Medicaid Services
- Purpose: A public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4000+ Medicare-certified hospitals (2015)

- Name of program and sponsor: Paul Coverdell National Acute Stroke Registry; Centers for Disease Control and Prevention
- Purpose: A national registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.
- Geographic area and number and percentage of accountable entities and patients included: 11 states; 403 hospitals (CDC, 2014)

- Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
- Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 hospitals (2015)

- Name of program and sponsor: Annual Report-Improving America's Hospitals; The Joint Commission
- Purpose: The annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

- Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint Commission
- Purpose: A certification program that recognizes that specific capabilities of hospitals that treat the most complex stroke cases.
- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

- Name of program and sponsor Disease-Specific Care Certification for Primary Stroke Centers; The Joint Commission
- Purpose: A certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
- Geographic area and number and percentage of accountable entities and patients included Nationwide; 1079 hospitals

- Name of program and sponsor Hospital Accreditation Program; The Joint Commission
- Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
- Geographic area and number and percentage of accountable entities and patients included Nationwide; 3300 Joint Commission-accredited hospitals (2014)

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not Applicable

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Not Applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

The rate of anticoagulation therapy for atrial fibrillation has steadily increased over the past five years based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 97.2%. A gap of ~12% still exists for the 10th percentile of hospitals. This trend is consistent with increases in the rate of anticoagulation for atrial fibrillation published by GWTG-Stroke and PCNASR. Non-Hispanic black patients continue to be an underserve population (Schwamm, 2010, Qian, 3013; CDC 2014). Women are also less likely to receive anticoagulation therapy for atrial fibrillation (Bushnell, et al, 2014).

- Geographic area and number and percentage of accountable entities and patients included: Nationwide; 1256 hospitals; 28,027 patients (The Joint Commission, 2014)

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not Applicable

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Novel oral anticoagulant drugs (NOACs) have been added to the list of comprehensive anticoagulant medications for stroke (Appendix C, Table 8,3) as they have become approved by the United States Food and Drug Administration (FDA). Although some adverse events have been associated with the NOACs and reported in the literature, e.g., increased incidence of myocardial infarction with dabigatran, no unintended negative consequences have been reported with the addition of the warfarin-alternative agents as a result of ongoing data collection.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0084 : Heart Failure (HF) : Warfarin Therapy Patients with Atrial Fibrillation
0241 : Stroke and Stroke Rehabilitation: Anticoagulant Therapy Prescribed for Atrial Fibrillation (AF) at Discharge
0624 : Atrial Fibrillation - Anticoagulation Therapy
1525 : Atrial Fibrillation and Atrial Flutter: Chronic Anticoagulation Therapy

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0624 : Atrial Fibrillation - Anticoagulation Therapy; Active Health Management – no longer NQF-endorsed.
0084 : Heart Failure (HF) : Warfarin Therapy Patients with Atrial Fibrillation –status unspecified; AMAPCPI
0241 : Stroke and Stroke Rehabilitation: Anticoagulant Therapy Prescribed for Atrial Fibrillation (AF) at Discharge- status unspecified; AAN

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Measure 1525 from the American College of Cardiology is a physician performance measure identified through CPT codes and could extend to the outpatient setting. The measure evaluates physician practice as opposed to hospital processes. The target population for measure 1525 differs from measure 0436 Anticoagulation Therapy for Atrial Fibrillation/Flutter in that it includes in the denominator population all patients age 18 years and older with a diagnosis of nonvalvular atrial fibrillation or atrial flutter whose assessment of the specified thromboembolic risk factors indicate one or more high-risk factors or more than one moderate risk factor, as determined by CHADS2 risk stratification. It is not specified for ischemic stroke patients with atrial fibrillation/flutter only.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not Applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Ann, Watt, awatt@jointcommission.org, 630-792-5944-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Karen, Kolbusz, kkolbusz@jointcommission.org, 630-792-5931-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel (TAP) is to provide advisory oversight in literature review, measure content and maintenance

of the specifications.
Members are:
The role of the Technical Advisory Panel (TAP) is to provide advisory oversight in literature review, measure content and maintenance of the specifications.
Members are:

Harold P. Adams, Jr., MD
University of Iowa Health Care
Iowa City, IA

Mark J. Alberts, MD
University of Texas Southwestern
Dallas, TX

Anne W. Alexandrov, RN
Health Outcomes Institute
Fountain Hills, AZ

Nadine Allyn RD
American Heart Association
Dallas, TX

Mary G. George, MD
Centers for Disease Control and Prevention
Atlanta, GA 30341

Martin Gizzi, MD
Legacy Health System
Portland, OR

Marianna Gorbaty
Mathematica Policy Research
Washington, DC

Judith Hinchey, MD
Tuft's Medical School
Steward St. Elizabeth's Medical Center
Boston, MA

Edward C. Jauch, MD
Medical University of South Carolina
Charleston, SC 2946

Irene Katzan, MD
Cleveland Clinic
Cleveland, OH

Kimberly E. Levasseur-Franklin, PharmD
Northwestern Memorial Hospital
Chicago, IL

Kathy Morrison, RN
Penn State Hershey Medical Center
Hershey, PA

Marilyn M. Rymer, MD
University of Kansas

Kansas City, MO

Jeffrey Saver, MD
UCLA Medical Center
Los Angeles, CA

Lee H. Schwamm, MD
Harvard Medical School
Massachusetts General Hospital
Boston, MA

Deborah Summers, RN
St Luke´s Brain and Stroke Institute
Kansas City, MO

Osama O. Zaidat, MD
St Vincent Mercy Hospital
Toledo, OH

Richard D. Zorowitz, MD
Medstar National Rehabilitation Network
Washington, DC

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2009
Ad.3 **Month and Year of most recent revision:** 10, 2015
Ad.4 **What is your frequency for review/update of this measure?** Biannual
Ad.5 **When is the next scheduled review/update for this measure?** 07, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0437
Measure Title: STK 04: Thrombolytic Therapy
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of acute ischemic stroke patients who arrive at this hospital within 2 hours of time last known well for whom IV t-PA was initiated at this hospital within 3 hours of time last known well. This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-6 Discharged on Statin Medication, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs. This measure captures the proportion of acute ischemic stroke patients who arrive at this hospital within 2 hours of time last known well for whom IV t-PA was initiated at this hospital within 3 hours of time last known well.

This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-6 Discharged on Statin Medication, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs.

Developer Rationale: Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Studies have demonstrated that the early administration of thrombolytic therapy within 3 hours of stroke symptom onset can significantly improve neurologic outcomes at 3 months in patients with ischemic stroke. The European Cooperative Acute Stroke Study (ECASS) III trial indicated that intravenous rtPA can be given safely to, and can improve outcomes for, carefully selected patients treated 3 to 4.5 hours after stroke; however, as the NINDS investigators concluded, the earlier that IV thrombolytic therapy is initiated, the better the patient outcome. Despite strong recommendations from the American Academy of Neurology, the American Heart Association, and the American College of Chest Surgeons, thrombolytic therapy is used in only a small proportion of ischemic stroke patients overall and in only a minority of eligible candidates.

Healthcare organizations that track IV thrombolytic administration for internal quality improvement purposes have seen a significant increase in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

Numerator Statement: Acute ischemic stroke patients for whom IV thrombolytic therapy was initiated at this hospital within 3 hours (less than or equal to 180 minutes) of time last known well.
Denominator Statement: Acute ischemic stroke patients whose time of arrival is within 2 hours (less than or equal to 120 minutes) of time last known well.
Denominator Exclusions:
- Less than 18 years of age
- Length of Stay > 120 days
- Enrolled in clinical trials related to stroke
- Admitted for elective carotid intervention
- Time last known well to arrival in the emergency department greater than 2 hours

• Documented reason for extending the initiation of IV thrombolytic
• Documented reason for not initiating IV thrombolytic

Measure Type: Process
Data Source: Electronic Clinical Data, Paper Medical Records
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2008 **Most Recent Endorsement Date:** Nov 01, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Summary of prior review in 2012:

- The body of evidence consistently supports that thrombolytic therapy results in a significant net reduction in the proportions of patients dead or dependent in activities of daily living due to ischemic stroke. Based on these findings, clinicians may choose to use thrombolytic therapy in selected patients within three hours of symptom onset.
- Class I, Level of Evidence A – 2007 AHA/ASA Guidelines for the Early Management of Adults with Ischemic Stroke, page 1676
    - Intravenous rtPA (0.9mg/kg, maximum dose 90 mg) is recommended for selected patients who may be treated within 3 hours of onset of ischemic stroke. Physicians should review the criteria outlined in Table 11 (which are modeled on those used in the NINDS trial) to determine the eligibility of the patient. A recommended regimen for observation and treatment of the patient is described in Table 12. This recommendation has not changed from previous statements.
- The 2012 Committee agreed that this measure was supported by strong evidence.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**
☐ **The developer provided updated evidence for this measure:**

Guidance from the Evidence Algorithm:
Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency high(Box 5) → Box 5a High

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and*

there is no need for repeat discussion and voting on Evidence?

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer provides the following national trend data:

	CY 2010	CY2011	CY 2012	CY 2013	CY 2014
# hospitals	129	136	136	224	1,099
# cases (den)	1,728	1,942	1,807	3,135	14,907
National aggregate rate	0.61	0.69	0.77	0.80	0.85
Mean hospital rate	0.57	0.57	0.67	0.75	0.75
50th percentile	0.67	0.72	0.77	0.88	0.89
10th and 90th percentiles	0.00 (10th) 1.00 (90th)	0.00 (10th) 1.00 (90th)	0.00 (10th) 1.00 (90th)	0.00 (10th) 1.00 (90th)	0.00 (10th) 1.00 (90th)

Disparities:

- The developer does not provide disparities data from use of this measure.
- The developer states that since the last submission, several large studies evaluating thrombolysis utilization have shown that while disparities still exist for many groups, the disparity gap has narrowed for certain groups, e.g., older adults, young adults, women, and young blacks.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Is there evidence that a gap in care still exits for minority populations? How can this measure be used to better understand disparities in care and outcomes for certain groups?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **No new studies/information

**High. Has not changed since we last reviewed this measure.

**High

1b. Performance Gap

Comments: **No disparities data provided

**Moderate. Some performance data is provided to show that the performance gap is reducing, but a sufficiently meaningful gap remains to justify the measure. The measurement does not appear to be topped out. Disparities data is more limited, in the form of several large studies which show some improvements in patient subgroups, but it is not clear from the measure how subgroup data is being captured to determine if disparities are being addressed.

**moderate – improving, disparities gap narrowing

1c. High Priority (previously referred to as High Impact)

Comments: **No comments provided

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): medical record abstraction (paper or electronic). This is not an eMeasure.

Specifications:

- Five data elements define the numerator and 14 data elements define the denominator.
- Developer reports the following revisions since last endorsement date:
  - A paragraph was added to the measure rationale referencing the European Cooperative Acute Stroke Study (ECASS) III trial and the safe administration of IV thrombolytic therapy within the extended timeframe of 3 to 4.5 hours after stroke for select patients, as well as, the importance of maintaining the administration target of 3 hours.
  - A new data element, Reason for Extending the Initiation of IV Thrombolytic, was added and the measure algorithm adjusted to exclude patients who receive IV thrombolytic therapy within 3 to 4.5 hours after the time last known well when a patient or medical reason is documented in the medical record. Notes for abstraction previously used to exclude these patients were removed from the data element IV Thrombolytic Initiation.
  - Abstraction guidelines for several data elements were also modified to address questions received from data abstractors. The timeframe for documentation of a Reason for Not Initiating IV Thrombolytic was specified as the day of or day after hospital arrival to prevent late entries and addendums after discharge. The same timeframe was specified for the new data element Reason for Extending the Initiation of IV Thrombolytic. Stand-alone reasons were specified for both reason data elements, and a new abstraction bullet added that system reasons are not acceptable.
  - For the data elements Date Last Known Well, Last Known Well, and Time Last Known Well, abstraction guidelines were clarified to address questions about multiple times or conflicting time documentation in the medical record. A sample of vendors and their hospital customers were surveyed to obtain feedback on these changes prior to implementation.
- Diagnosis and procedure codes have been converted from ICD-9 to ICD-10-CM diagnosis and ICD-10-PCS procedure codes. See attached spreadsheet Appendix A Table 8.1 (ischemic stroke).
- A conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications:
  - "The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015 developer advisory panel determined that the intent of the measure was changed by the conversion to ICD-10 and a crosswalk was posted."
- Sampling monthly or quarterly is allowed; instructions for sampling are provided.
- A web-based data collection tool has been developed but is not described. Hospitals are not required to use this tool.
- The measures is specified at the hospital level of analysis.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Inter rater-reliability of the data elements for one year (4Q2010-3Q2011) in 77 hospitals and 739 patient records showed an overall agreement rate of 98.1%.
  - Five data elements found >90% agreement: Clinical Trial (99.7%), ED Patient (99.7%), Elective Carotid Intervention (99.6%), IV Thrombolytic Initiation (90.7%), and IV Thrombolytic Initiation Date (92.6%).
  - Four data elements found >80% agreement: Date Last Known Well (87.9%), IV Thrombolytic Initiation Time (85.2%), Last Well Known (87.2%), and Time Last Known Well (80.0%).
  - One data element found <80% agreement: Reason for Not Initiating IV Thrombolytic Therapy (76.7%).
  - No Kappa scores were presented.

- In 2012, the Committee requested further information from the developers on why patients with a length of stay over 120 days are excluded from the measure. The developer explained that this exclusion is an artifact of the billing cycles used by CMS, and that in practice, it has a negligible impact on the measure.

Describe any updates to testing: No new information provided

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing: see above

Results of reliability testing: see above

Guidance from the Reliability Algorithm:

Precise specifications (Box 1) →empirical testing as specified (Box 2) → reliability testing with patient-level data elements (Box 8) → assessed all critical data elements (Box 9)→ high/moderate certainty the data elements are reliable (Box 10) → Moderate (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Questions for the Committee:

- *The developer has not provided any new reliability testing. Does the Committee agree there is no need for repeat discussion and voting on Reliability?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the

evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

Question for the Committee:

o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Face validity of the data was assessed via survey and focus groups of hospitals participating in the pilot test.
- The developer reports that "All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant."
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.
- Ongoing feedback from measure users that is systematically reviewed to identify trends and/or measure specifications that require clarification or revision. Predominantly, questions regarding clarification of the data elements Last Known Well, Date Last Known Well, and Time Last Known Well were most common.

Describe any updates to validity testing – New empirical validity testing data is presented

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- Pearson Correlation Coefficients were calculated for seven stroke measures for hospitals that reported 12 months of data and had 30 or more denominator cases for the year (2014-2015).Presumably high performing hospitals would do well on all stroke performance measures.

Validity testing results:

- Analysis of 1,318 hospitals and 2,206,379 patients records generated a table of Pearson Correlation Coefficient results that shows a statistically significant (P<.0001), positive correlation of this measure with six other stroke performance measures supporting the hypothesis that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

Questions for the Committee:

- o *Is the test sample adequate to generalize for widespread implementation?*
- o *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- o *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients are excluded from the measure for the following reasons:

- Patients less than 18 years of age
- Patient who have a Length of Stay greater than 120 days
- Patient enrolled in a Clinical Trial
- Patients admitted for Elective Carotid Intervention
- Time Last Known Well to arrival in the emergency department greater than 2 hours
- Patients with a documented Reason For Extending the Initiation of IV Thrombolytic
- Patients with a documented Reason For Not Initiating IV Thrombolytic

- New data on the frequency of exclusions is presented for 2, 206,379 admissions in 1,318 hospitals:
  - Clinical Trial: 0.25% (range 0.24-10.0%)
  - Elective Carotid Intervention: 11.8% (range 0.16-95.2%)
  - Time Last Known Well: None
  - Reason For Not Initiating IV Thrombolytic: 24.0% (range 0.42-92.0%)
  - Reason For Not Initiating IV Thrombolytic: 0.70% (range 0.09-35.2%)

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Performance data is presented above under opportunity for improvement. Complete details of the data is presented in 1b.
- The developer explains their approach: "There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations."

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods: Not Applicable

2b7. Missing Data

- The developer describes the handling of missing data in the measure specifications (S.22) but does not provide information on the frequency of missing data or potential impact on measure results.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → potential threats to validity mostly assessed (Box2) →empirical validity testing (Box 3) → measure score testing (Box 6) → sound conceptualization and hypotheses (Box 7) →high confidence that score are a valid indicator of quality(Box 8a) → High

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **Consistent

**The specification is consistent with the evidence.

**sufficient validity

2a2. Reliability Testing

Comments: **Sufficient validity testing

**New data demonstrating a highly significant positive correlation between this measure and 6 other stroke performance measures. This is in addition to the previous validity data provided in the last review. The validity data in the last endorsement was deemed valid and remains so. This new data demonstrates a correlation among the various stroke measures, suggesting that increasing quality on this measure is associated with similar levels of quality on other measures on a per hospital basis. If there is sufficient data to support the validity of the other stroke measures as a quality measure, then this new data does support the validity of this measure. However, without independent data on the validity of those other measures, it is not clear how the positive correlation alone further supports the validity of this measure.

**yes, data strongly supports

2b2. Validity Testing

Comments: **No

**Overall the exclusions are consistent with the evidence. However, it is not clear if simple documentation of a reason for extending or not initiating IV thrombolytics is sufficient. For example, would documentation of an unjustified reason be acceptable here? There does not appear to be a criteria that the documented reason be supported by evidence. For example, if there was a delay in IV thrombolysis to the 3-4.5 hour window for a reason that is not supported by the ECASS study, but this was documented, would that be sufficient to exclude from the denominator?

**no

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **New reliability testing not needed

**The developer has not provided new reliability data since the last review. There is no compelling reason for repeat discussion and vote, although it seems possible that one element that was below 80%, Reason for not initiating IV thrombolysis, could be affected by or confounded by the new data element derived from the ECASS study regarding the 3-4.5 hour window.

**yes, widely accepted

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is medical record abstraction: not all hospitals are able to generate the data electronically so a paper-based option is available.
- Some data elements are in electronic form.
- Data collection burden is significant.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Some data elements in EHR

**Most of the data elements are routinely generated. The are not available in electronic form in institutions without an EHR so a paper option is provided in this measure.

**High fasibility, with caveat of limitations for institutions without HER. However, this is generally an ER or ICU procedure and there

is high likelihood of electronic data capture in these settigns for all institutions.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- The Joint Commission Quality Check®: Public website that allows consumers to search for accredited and certified organizations.
- CMS Hospital Compare: Public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.
- CDC Paul Coverdell National Acute Stroke Registry: National registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.
- CMS Hospital Inpatient Quality Reporting Program (IQR): The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- The Joint Commission Annual Report-Improving America's Hospitals: Annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
- The Joint Commission Disease-Specific Care Certification for Comprehensive Stroke Centers: Certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
- The Joint Commission Disease-Specific Care Certification for Primary Stroke Centers: Certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results:

- The developer provided the following progress information on 1,099 hospitals nationwide and 14,907 patients:
    - The rate of intravenous thrombolytic (t-PA) administration has improved by 57.8% over the past five years based on Joint Commission ORYX performance measurement data. This trend is consistent with increases in the rate of IV t-PA administration reported in multiple published studies.
    - Other trends identified in the literature include: an increasing interest in the concept of treating patients with milder symptoms or rapid improvement; an increased proportion of very old (> 85 years) acute ischemic stroke patients receiving tPA; an increased proportion of young acute ischemic stroke patients (mean age 37 years), in particular young blacks, receiving tPA; and, a slight increase in the proportion of women receiving tPA. Additionally, an increased trend in the rate of tPA utilization for urban patients treated at urban hospitals, as well as, rural patients treated at rural hospitals has been reported.

Unexpected findings (positive or negative) during implementation: The developer did not identify any unexpected findings during implementation.

Potential harms:

- The updated guideline recommendation for IV t-PA administration within 3 to 4.5 hours after stroke generated concern that thrombolytic therapy initiation would be delayed for patients eligible to receive treatment within 3 hours and result in poorer outcomes for stroke patients. Since earlier initiation is associated with better neurological outcomes, the STK-4 numerator has remained unchanged; however, the measure specifications have been modified to clearly exclude patients who are treated with IV t-PA within 3 to 4.5 hours and have a documented medical or patient reason for the delay in initiation.

Feedback :

- The developer states that feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As previously noted, measure users have indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences
Comments: **High usability
**The benefits clearly outweigh any potential unintended consequences. However, it is not entirely clear if simple exclusion of patients with a documented medical or patient reason for receiving delayed tPA (3-4.5 hours) is adequate. A documented reason could be inappropriate, due to physician ignorance of current data, but it would still be documented.
**High usability. No anticipated unintended consequences from measurement.

Criterion 5: Related and Competing Measures

Related or competing measures:

- 0164 : Fibrinolytic Therapy received within 30 minutes of hospital arrival
- 0288 : Fibrinolytic Therapy Received Within 30 Minutes of ED Arrival
- 1952 : Time to Intravenous Thrombolytic Therapy

- 0242 : Stroke and Stroke Rehabilitation: Tissue Plasminogen Activator (t-PA) Considered – AMA-PCPI – no longer NQF endorsed

Harmonization:

- Measures 0288 and 0164 are AMI (Acute Myocardial Infarction) measures and are harmonized to the extent that the measures utilize some of the same data elements.
- The target population for measure 1952 from the American Heart Association/American Stroke Association also includes patients hospitalized for acute ischemic stroke; however, the measure captures average door-to-needle time and uses a target of less than 60 minutes rather than the proportion of patients who arrive within 2 hours

and receive t-PA within 3 hours of time last known well.

- Measure 0242 is a physician level measure with a targeted population of ischemic stroke patients identified through CPT codes and could extend to the outpatient setting.

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0437	NQF Project: Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT
Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria)
1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*: The focus of the measure is to improve neurological outcomes for ischemic stroke patients through the timely initiation of intravenous (IV) thrombolytic therapy (t-PA) to carefully screened, eligible candidates. IV thrombolytic therapy initiated >> improved neurological outcomes >> decreased morbidity and mortality. 1c.2-3 Type of Evidence *(Check all that apply)*: Clinical Practice Guideline Selected individual studies (rather than entire body of evidence) Systematic review of body of evidence (other than within guideline development) 1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*: This measure is consistent with the guidelines recommended by the American Heart Association/American Stroke Association for the early management of adults with ischemic stroke. The administration of thrombolytic agents to carefully screened, eligible patients with acute ischemic stroke has been shown to be beneficial in several clinical trials. These included two positive randomized controlled trials in the United States: The National Institute of Neurological Disorders and Stroke (NINDS) Studies, Part I and Part II. Based on the results of these studies, the Food and Drug Administration approved the use of intravenous recombinant tissue plasminogen activator (IV r-TPA or t-PA) for the treatment of acute ischemic stroke when given within 3 hours of stroke symptom onset. A large meta-analysis controlling for factors associated with stroke outcome confirmed the benefit of IV t-PA in patients treated within 3 hours of symptom onset. While controversy still exists among some specialists, the major society practice guidelines developed in the United States all recommend the use of IV t-PA for eligible patients. Physicians with experience and skill in stroke management and the interpretation of CT scans should supervise treatment. 1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: Relevant literature supporting thrombolysis for acute ischemic stroke was identified through a Cochrane review, prepared and maintained by The Cochrane Collaboration and published in The Cochrane Library 2009, Issue 4. The search strategy employed the Cochrane Stroke Group Trials Register (last

searched October 2008), MEDLINE (1966 to October 2008) and EMBASE (1980 to October 2008). Selection criteria involved randomized trials of any thrombolytic agent compared with control in patients with definite ischemic stroke (i.e., computerized tomography (CT) or magnetic resonance imaging (MRI) having excluded intracranial hemorrhage prior to randomization).

The review included 26 trials involving 7152 patients. This review included all new trials completed and made public since 2003, as well as additional data published since 2003 from trials included in earlier versions of the review. The total number of patients included in the review represented a 10-fold increase in the number of study patients since the first review in 1990. Very few of the patients (0.5%) were aged over 80 years.

The primary outcomes measured were death or dependency, as defined by modified Rankin score of 3 to 6, and death at the end of follow-up. Not all trials contributed data to each outcome. The trials tested urokinase, streptokinase, recombinant tissue plasminogen activator, recombinant pro-urokinase or desmoteplase, Four trials used intra-arterial administration, the rest used the intravenous route. Most data came from trials that started treatment up to six hours after stroke; three trials started treatment up to nine hours and one small trial up to 24 hours after stroke. About 55% of data (patients and trials) came from trials testing intravenous tissue plasminogen activator.

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting thrombolysis for acute ischemic stroke is high. The Cochrane Review noted above confirmed that thrombolytic treatment can reduce the risk of disability post-ischemic stroke, despite the bleeding risks. This review was conducted by two review authors who applied the selection criteria and extracted data. The reviewers assessed trial quality and verified the extracted data with the principal investigators of all major trials for both published and unpublished data. Specifically, the reviewers hand-searched journals from 1979 to April 1994; contacted 321 pharmaceutical companies for more information about trials known to exist and unknown based on prior searches; examined references quoted in thrombolytic therapy papers; contacted principal investigators of trials in Europe, North American, Japan, China , and Australia; and, attended multiple international conferences on stroke and thrombolysis since 1991). Every effort was made to identify truly randomized trials of thrombolytic therapy compared with placebo or open control in patients with acute ischemic stroke. Trials that were not truly random, such as dose-range-finding studies, were excluded.

According to the Cochrane reviewers, many trials had some imbalance in key prognostic variables. Several trials did not have complete blinding of outcome assessment. Thrombolytic therapy administered up to six hours after ischemic stroke significantly reduced the proportion of patients who were dead or dependent (modified Rankin 3 to 6) at three to six months after stroke (odds ratio (OR) 0.81, 95% Confidence Interval (CI) 0.73 to 0.90). Thrombolytic therapy increased the risk of symptomatic hemorrhage (OR 3.49, 95% CI 2.81 to 4.33) and death within three to six months after stroke (OR 1.31, 95% CI 1.14 to 1.50). Treatment within three hours of stroke appeared more effective in reducing death or dependency (OR 0.71, 95% CI 0.52 to 0.96) with no statistically significant adverse effect on death (OR 1.13, 95% CI 0.86 to 1.48). There was heterogeneity among the trials in part attributable to concomitant antithrombotic drug use (P =0.02), stroke severity and time to treatment. Antithrombotic drugs given soon after thrombolysis may increase the risk of death.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of evidence consistently supports that thrombolytic therapy results in a significant net reduction in the proportions of patients dead or dependent in activities of daily living due to ischemic stroke. Based on these findings, clinicians may choose to use thrombolytic therapy in selected patients within three hours of symptom onset.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
There is strong evidence on the immediate hazards and net benefits of thrombolytic therapy in the treatment of acute ischemic stroke. Thrombolytic therapy administered up to six hours after ischemic stroke significantly reduced the proportion of patients who were dead or dependent (modified Rankin 3 to 6) at three to six months after stroke (odds ratio (OR) 0.81, 95% Confidence Interval (CI) 0.73 to 0.90). Treatment within three hours of stroke appeared more effective in reducing death or dependency (OR 0.71, 95% CI 0.52 to 0.96) with no statistically significant adverse effect or death (OR 1.13, 95% CI 0.86 to 1.48).

Treating ischemic stroke patients with IV t-PA within 3 hours of time last known well improves functional outcomes at 3 months. Considering the mean lifetime cost of ischemic stroke, estimated at $140,048 per person, thrombolytic therapy is also cost-effective.

Fagan et al. (1998) used data from the NINDS rt-PA Stroke Trial and developed a Markov model to compare costs per 1,000 eligible t-PA patients with the costs per 1,000 untreated patients. This analysis revealed a greater than 90% probability of costs savings when eligible patients are treated with t-PA within 3 hours. In the NINDS rt-PA Stroke Trial, the average length of stay was significantly shorter for patients treated with t-PA than the placebo group (10.9 versus 12.4 days; p=0.002), and more patients were discharged to home than to an inpatient rehabilitation facility or a nursing home (48% versus 36%; p=0.002). The Markov model estimated an increase in hospitalization costs of $1.7 million and a decrease in rehabilitation costs of $1.4 million and nursing home cost of $4.8 million per 1,0000 eligible t-PA patients. Furthermore, the analysis estimated the long-term impact on health outcomes to be 564 quality-adjusted life-years saved over 30 years per 1,000 patients.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: American Heart Association. During our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, and provided information that was accessible and met the requirements set out in this measure maintenance form.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.13 Grade Assigned to the Body of Evidence: Class I, Level of Evidence A

1c.14 Summary of Controversy/Contradictory Evidence: The benefit of IV thrombolytic therapy within 3 hours of stroke symptom onset is undisputed. No position against the administration of thrombolytic therapy to select, eligible ischemic stroke patients was noted in the literature. However, there is still not enough evidence to answer some questions. More randomized trials are needed to determine the full impact of benefit, the latest time window for safe administration, the cut-off age for treatment, the type of stroke and grades of severity most likely to respond favorably to treatment, and the impact of various co-morbidities and concomitant drug therapies on outcome.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- del Zoppo GJ, Saver JL, Jauch EC, Adams HP. Expansion of the Time Window for Treatment of Acute Ischemic Stroke With Intravenous Tissue Plasminogen Activator: A Science Advisory From the American Heart Association/ American Stroke Association. Stroke. 2009;40:2945-2948.
- Diagnosis and Initial Treatment of Ischemic Stroke, Institute for Clinical Systems Improvement (ICSI), 2001.
- Fagan SC, Morgenstern LB, Petitta A, Ward RE, Tilley BC, Marler JR, Levine SR, Broderick JP, Kwiatkowski TG, Frankel M, Brott TG, Walker MD, and The NINDS rt-PA Stroke Study Group. Cost-efffectiveness of tissue plasminogen activator for acute ischemic stroke. Neurology. 1998;50(4):883-890.
- Guyatt GH, Akl EA, Crowther M, Gutterman DD, Schu?nemann HJ, and for the American College of Chest Physicians Antithrombotic Therapy and Prevention of Thrombosis Panel. Executive summary: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141: 33S.
- Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Gidetti D, et. al. Thrombolysis with Alteplase 3 to 4.5 hours after acute ischemic stroke. The European Cooperative Acute Stroke Study (ECASS) Investigators. NEJM. 2008;359(13):1317-29.
- Hacke W, Kaste M, Fieschi C, et al. Intravenous thrombolysis with recombinant tissue plasminogen activator for acute hemispheric stroke. The European Cooperative Acute Stroke Study (ECASS). JAMA 1995;274:1017-1025.
- Management of Patients with Stroke. Assessment, investigation, immediate management and secondary prevention, Scottish Intercollegiate Guidelines Network, 1997.
- Marler JR, Tilley BC, Lu M, Brott TG, Lyden PC, Grotta JC, Broderick JP, Levine SR, Frankel MP, Horowitz SH, Haley EC, Lewandowski CA, Kwiatkowski TP. Early stroke treatment associated with better outcome The NINDS rt-PA Stroke Study. Neurology 2000;55: 1649-1655.
- Sacco RL, Adams R, Albers G, Alberts MJ, Benavente O, Furie K, Goldstein LB, Gorelick P, Halperin J, Harbaugh R, Johnston SC, Katzan I, Kelly-Hayes M, Kenton EJ, Marks M, Schwamm LH, Tomsick T. Guidelines for Prevention of Stroke in Patients With Ischemic Stroke or Transient Ischemic Attack: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association Council on Stroke: Co-Sponsored by the Council on Cardiovascular Radiology and Intervention. Stroke. Vol. 37, 2006:577.
- STROKE the First Hours Guidelines for Acute Treatment, National Stroke Association, 2000.
- The ATLANTIS, ECASS, and NINDS rt-PA Study Group Investigators. Association of Outcome with early stroke treatment:

pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke Trials. Lancet 2004;363:768-774.
- The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. New England Journal of Medicine 1995;333:1581-1587.
- Wardlaw JM, Murray V, Berge E, del Zoppo GJ for The Cochrane Collaboration. Thrombolysis for acue ischaemic stroke (review). The Cochrane Library. 2009:4: 1-131.

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)*:
2007 AHA/ASA Guidelines for the Early Management of Adults with Ischemic Stroke, page 1676.

Class I, Level of Evidence A
Intravenous rtPA (0.9mg/kg, maximum dose 90 mg) is recommended for selected patients who may be treated within 3 hours of onset of ischemic stroke. Physicians should review the criteria outlined in Table 11 (which are modeled on those used in the NINDS trial) to determine the eligibility of the patient. A recommended regimen for observation and treatment of the patient is described in Table 12. This recommendation has not changed from previous statements.

1c.17 Clinical Practice Guideline Citation: Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A Guideline From the American Heart Association/American Stroke, Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1674-1677.

1c.18 National Guideline Clearinghouse or other URL:
http://www.guidelines.gov/content.aspx?id=10911&search=early+management+of+adults+with+ischemic+stroke

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: This guideline was approved by the American Heart Association Science Advisory and Coordinating Committee on January 6, 2007. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.23 Grade Assigned to the Recommendation: Class I, Level of Evidence A

1c.24 Rationale for Using this Guideline Over Others: The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: High 1c.26 Quality: High1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0437_Evidence_MSF5.0_Data.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Studies have demonstrated that the early administration of thrombolytic therapy within 3 hours of stroke symptom onset can significantly improve neurologic outcomes at 3 months in patients with ischemic stroke. The European Cooperative Acute Stroke Study (ECASS) III trial indicated that intravenous rtPA can be given safely to, and can improve outcomes for, carefully selected patients treated 3 to 4.5 hours after stroke; however, as the NINDS investigators concluded, the earlier that IV thrombolytic therapy is initiated, the better the patient outcome. Despite strong recommendations from the American Academy of Neurology, the American Heart Association, and the American College of Chest Surgeons, thrombolytic therapy is used in only a small proportion of ischemic stroke patients overall and in only a minority of eligible candidates.

Healthcare organizations that track IV thrombolytic administration for internal quality improvement purposes have seen a significant increase in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Many patients with acute ischemic stroke arrive at the hospital within three hours of stroke onset without documented contraindications who still do not receive intravenous (IV) thrombolysis (t-PA). Although rates of IV t-PA administration have improved over time, there is still less than optimal performance especially in the lower quartile and decile of hospitals. In October, 2009, The Joint Commission added the stroke (STK) measure set as a new core measure option to meet performance measure requirements for Joint Commission hospital accreditation purposes. Joint Commission ORYX performance measure data for 4Q 2009 yielded an average measure rate of 48.9% from 39 hospitals collecting data for this measure (n=233 patients). The average rate for all hospitals collecting data for this measure (i.e., 1099 hospitals; n=14,907 patients) is currently 84.5%, indicating that a potential performance gap of 15% persists if the optimal rate is 100%. Below is the specified level of analysis for STK-4 beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 233 denominator cases; 114 numerator cases; 39 hospitals; 0.48927 national aggregate rate; 0.54417 mean of hospital rates; 0.41298 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.625 50th percentile rate/median rate; 0.03704 25th percentile rate/lower quartile; and, zero 10th percentile rate.

CY 2010: 1728 denominator cases; 1053 numerator cases; 129 hospitals; 0.60938 national aggregate rate; 0.56977 mean of hospital rates; 0.37503 standard deviation; 1.0 90th percentile rate; 0.9 75th percentile rate/upper quartile; 0.66667 50th percentile rate/median rate; 0.21212 25th percentile rate/lower quartile; and, zero 10th percentile rate.

CY 2011: 1942 denominator cases; 1333 numerator cases; 136 hospitals; 0.68641 national aggregate rate; 0.57338 mean of hospital rates; 0.37267 standard deviation; 1.0 90th percentile rate; 0.89737 75th percentile rate/upper quartile; 0.71964 50th percentile rate/median rate; 0.23611 25th percentile rate/lower quartile; and, zero 10th percentile rate.

CY 2012: 1807 denominator cases; 1393 numerator cases; 136 hospitals; 0.77089 national aggregate rate; 0.67283 mean of hospital rates; 0.34402 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.7735 50th percentile rate/median rate; 0.5 25th percentile rate/lower quartile; and, zero 10th percentile rate.

CY 2013: 3135 denominator cases; 2493 numerator cases; 224 hospitals; 0.79522 national aggregate rate; 0.74944 mean of hospital rates; 0.3186 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.875 50th percentile rate/median rate; 0.66667 25th percentile rate/lower quartile; and, zero 10th percentile rate.

CY 2014: 14,907 denominator cases; 12,598 numerator cases; 1099 hospitals; 0.84511 national aggregate rate; 0.75482 mean of hospital rates; 0.32814 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.88889 50th percentile rate/median rate; 0.66667 25th percentile rate/lower quartile; and, zero 10th percentile rate.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
Not applicable.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
Since the last endorsement period, several large studies have evaluated trends in thrombolysis utilization across the United States. While disparities still exist for many groups, the disparity gap has narrowed for certain groups, e.g., older adults, young adults, women, and young blacks, since the last submission.

George and colleagues (2015) analyzed discharge data from 2005 to 2010 in the Nationwide Inpatient Sample (NIS) of the Agency for Healthcare Research and Quality (AHRQ), Healthcare Cost and Utilization Project (Rockville, MD). This analysis identified persistent racial disparities for blacks and Hispanics consistent with those found previously for blacks and Hispanics (Kimball, et al. 2012), as well as, a trend to treat older adults who would not have been considered t-PA candidates in earlier years and increased utilization rates for rural patients. From this database, 512,429 older adults (> 65 years of age) with acute ischemic stroke who received an injection or infusion of a thrombolytic agent were identified and divided into three age groupings (65-74, 75-84, > 85 years). Among U.S. hospitals with acute stroke patients, there were an estimated 1462 hospitals (32%) administering thrombolysis in 2005 and 1884 hospitals (43%) in 2010 (trend P = 0.003). For older adult stroke admissions, the rate of thrombolysis increased from 1.7% (95% CI: 1.6-1.8%) in 2005 to 5.4% (95% CI: 5.2-5.6%) in 2010, representing a three-fold increase in thrombolysis rates for older adults (trend P < 0.001). The largest increases occurred for individuals > 85 years of age with an approximate four-fold rate increase from 2005 to 2010. Rates of administration increased three-fold for urban patients and urban hospitals. Thrombolysis rates also increased for rural patients (0.9% in 2005 vs. 3.3% in 2010; trend P < 0.001), and rural hospitals increased at a slower rate (0.5% in 2005 vs. 1.7% in 2010; (trend P < 0.001). Low volume hospitals increased their rates of thrombolysis in older adults to a lesser degree than higher volume centers.

A second study using discharge data obtained from the Nationwide Inpatient Sample between 2001 and 2009 reported that disparities for young blacks has significantly improved in recent years (Kansara, et al. 2013). Between 2001 and 2009, there were an estimated 4,917, 217 admissions for acute ischemic stroke. Of these, 204,703 (4.16%) were young patients with a mean age approximately 37 years. The use of thrombolysis for young acute ischemic stroke patients increased 270% during this time period. The increased rate was noted across all races, including white, black, and nonwhite/nonblack populations. Unlike previous studies that reported that black patients were less likely to receive thrombolysis, this study found that a greater percentage of young black patients with acute ischemic stroke (5.45%) received thrombolysis than young white patients with acute ischemic stroke (4.57%) in 2009.

A univariate analysis of more than one million (N=1,093,895) acute ischemic stroke patients from 1683 hospitals participating in the American Heart Association's Get With the Guidelines-Stroke database was conducted to evaluate changes in the patterns of IV t-PA use over the 9-year period from April 2003 to December 2011. IV thrombolytic use has changed over time with a broader range of patients treated in later years. According to this analysis, the proportion of patients age > 85 years treated with IV t-PA increased

from 10.5% in 2003-2005 to 16.4% in 2010-2011 (P<0.001). Also, the gender distribution of t-PA use changed slightly, with the proportion of t-PA use among women increasing from 48.6% to 51%. The population receiving t-PA also became more diverse, with nonwhites accounting for 21.1% of t-PA use in 2003 to 2005 but 28.9% in 2010 to 2011(P<0.001) (Schwamm, et al. 2013). This study is among the first to describe the temporal trends of the past decade in IV t-PA use in patients with acute ischemic stroke in a clinically derived dataset from a sizeable cohort of U.S. hospitals nationwide.

Another published study utilized the American Heart Association's Get With the Guidelines-Stroke registry linked with Medicare claims data set to examine whether 30-day and 1-year outcomes differed by race/ethnicity among older patients with acute ischemic stroke (Qian, et al., 2013). Compared with other race/ethnicity groups, non-Hispanic black patients were less likely to receive IV t-PA in less than 3 hours from stroke onset. Relative to whites, black and Hispanic patients had higher adjusted 1-year all-cause rehospitalization (black: adjusted odds ratio, 1.28 [95% CI, 1.21-1.37]; Hispanics: adjusted odds ratio 1.22 [95% CI, 1.11-1.35]. Non-Hispanic black patients were more likely to be treated at high-volume and academic hospitals, which were generally located in the south. These findings were based on an analysis of 200,900 patients, including 20514 non-Hispanic blacks (10.2%), with acute ischemic stroke greater than 65 years of age from 926 U.S. centers participating in the GWTG-Stroke program from April 2003 through December 2008.

A prior 2011 report from the American Heart Association/American Stroke Association reported that racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al. 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

As stated in the previous submission citing earlier studies, the burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

In the national REGARDS cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years of age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age : RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined

significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence
rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer,et al., 2010). In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).

Minorities are less likely to use emergency medical services and more likely to wait longer before going to the hospital than are non-Hispanic whites. After hospital arrival, blacks or African Americans and possibly Hispanics experience longer wait times in the emergency department which results in treatment delays that include administration of thrombolytic therapy. Although unmeasured factors may play a role in these delays, the presence of bias in the delivery of care cannot be excluded (Cruz-Flores, et al, 2011).

The Department of Neurology and Stroke Program, Wayne State University School of Medicine, Detroit, MI (Bhattacharya P, et al., 2011) explored racial differences in the delivery of care to patients with acute stroke between hospitals certified in primary stroke care by The Joint Commission and noncertified hospitals. A retrospective chart review of 574 patients (25.1% African American) with ischemic stroke admitted to five Joint Commission certified primary stroke centers and five non-certified hospitals was conducted. Similar to previous studies, Bhattacharya found that African Americans often did not receive intravenous tPA because of a delay in hospital arrival. Whites were more likely to arrive by emergency transport services (65.5% vs 51.1%; P= 0.004) to be evaluated by a stroke team (19.1% vs. 7.7%; P=0.001), and to have documented National Institutes of Health Stroke Scale (NIHSS) score (40.2% vs. 29/9%; P=0.03); however, the number of white and black patients who received IV t-PA was not statistically different (2.1% in African Americans, 3.5% in Caucasians; P=0.40).

A larger study of 1044 patients (74% African American, 19% non-Hispanic white) with ischemic stroke (Hsia AW, et al., 2011), found that blacks were one-third less likely than whites to receive IV t-PA (3% vs.10%, P<0.001). Blacks were less likely than whites to present in 3 hours of symptom onset (13% vs. 21%; P=0.004). They were also less likely to be eligible candidates for thrombolytic therapy (5% vs. 13%; P<0.001). Of those patients who presented in 3 hours, blacks were almost half as likely to be treated with IV t-PA when compared to whites (27% vs. 46%; P=0.023).

A recent study from the Mayo Clinic, Rochester, MN (Naser DM, et al., 2011) investigated possible racial and ethnic disparities in the administration and outcome of recombinant tissue plasminogen activator (rt-PA) therapy for acute ischemic stroke in whites, blacks, Hispanics, and Asian/Pacific Islanders. Patients with a primary diagnosis of acute ischemic stroke who received rt-PA were identified using data from the National Inpatient Sample for 2001-2008 and stratified by race. The investigators analyzed the association of patient race on rt-PA utilization rate, in-hospital morbidity (i.e., discharges to a long-term care facility), intracranial hemorrhage (ICH) rate, and in-hospital mortality. Multivariate logistic regression analysis was performed to identify independent predictors of poor outcomes. Naser and colleagues concluded that whites had a higher rate of t-PA utilization than black and Hispanic patients (2.3% vs. 2.2% P= 0.07), although not statistically significant. Multivariate analysis of morbidity, mortality and ICH rates found that Asian/Pacific Islanders had significantly higher rates of mortality (odds ratio, 1.22, 95% CI, 1.91-2.11; P < .0001) compared with whites. Thrombolytic utilization was greater in white and Asian/Pacific Islander patients than in black and Hispanic patients. Asian/Pacific Islander race was associated with increased risk of ICH and mortality and rt-PA administration.

• Bhattacharya P, Mada F, Salowich-Palm L, et al. Department of neurology and Stroke Program, Wayne State University School of Medicine, Detroit, Michigan. Are racial disparities in stroke care still prevalent in certified stroke centers? J Stroke Cerebrovasc Dis. 2011.
• Brown AF, Liang LJ, Vassar SD, et al. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-8.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, et al, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• George BP, Asemota AO, Dorsey ER, Haider AH, Smart BJ, Urrutia VC, Schneider EB. United States trends in thrombolysis for older

adults with acute ischemic stroke. Clinical Neurology and Neurosurgery. 2015;139:16-23.
• Gillum RF, Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 2011;42(11):3294-6.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-75.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Hsia AW, Edwards DF, Morgenstern LB, et al. Racial disparities in tissue plasminogen activator treatment rate for stroke: a population-based study. Stroke. 2011;42(8):2217-21.
• Jacobs BS, Birbeck G, Mullard AJ, Hickenbottom S, Kothari R, Roberts S, Reeves MJ. Quality of hospital care in African Americans and white patients with ischemic stroke and TIA. Neurology.2006;66:809-14.
• Johnston SC, Fung LH, Gillum LA, Smith WS, Brass LM, Lichtman JH, Brown AN. Utilization of intravenous tissue-type plasminogen activator for ischemic stroke at academic medical centers.: the influence of ethnicity. Stroke. 2001; 32:1061-68.
• Kansara A, Chaturvedi S, Bhattacharya P. Thrombolysis and outcome of young stroke patients over the last decade: insights from the nationwide inpatient sample. J. Stroke Cerebrovasc Dis. 2013;6:799-804.
• Karve S, Balkrishnan R, Seiber E Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Morris DL, Rosamond W, Madden K, Schultz C, Hamilton S. Pre-hospital and emergency department delay after acute stroke: the Genetech Stroke Presentation Survey. Stroke. 2000;31:2585-90.
• Kimball MM, Neal D, Waters MF, Hoh BL, Race andincome disparity in ischemic stroke care: nationwide inpatient sample database, 2002-2008. Cerebrovasc Dis. 2012.
• Kissela BM, Khoury JC, Alwell K, et al. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, et al. Stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Morgenstern LB, Smith MA, Sánchez BN, et al. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol.
2013;74:778–785.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Nasar DM, Brinjikji W, Cloft HJ, Rabinstein AA. Racial and ethnic disparities in the use of intravenous recombinant tissue plasminogen activator and outcomes for acute ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, et al. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e-78-e82.
• Schwamm LH, Syed FA, Reeves MJ, Smith EE, Saver JL, Messe S, Bhatt DL, Grau-Sepulveda MV, Peterson ED, Fonarow GC. Temporal trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With the Guidelines-Stroke hospitals. Circ Cardiovasc Quality Outcomes. 2013;6:543-549.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality, Severity of illness

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States remains attributable to stroke. More women than men die of stroke each year. Women account for almost 60% of US stroke deaths. (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, et al., 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 ($P<0.05$)(US Burden of Disease Collaborators, 2013). Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

Thrombolytic therapy is one of the most promising treatments for acute ischemic stroke. The majority of strokes are due to blockage of an artery in the brain by a blood clot. Prompt treatment with clot dissolving (thrombolytic) drugs can restore blood flow before major brain damage has occurred. In the United States, Canada, and most European countries, alteplase/recombinant tissue plasminogen activator (rt-PA) has been approved for use within three hours of stroke symptom onset. Successful treatment is likely to improve neurological outcomes for ischemic stroke patients at three months and later; however, intracranial hemorrhage is a serious complication of therapy and may be fatal (Jauch EC, et al., 2013).

Clinical practice guidelines for intravenous thrombolysis with rt-PA (Adam HP, et al., 2007) cite the National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Study (1995), which partially supported the approval of rt-PA by the U.S. Food and Drug Administration (FDA). The NINDS trial was conducted in two consecutive parts. Part A trials in the 1980s studied very low doses of thrombolytic therapy, given daily by intravenous route for several days. The onset of treatment occurred anytime from 5 to 14 days post symptom onset. Part A trials did not collect data on functional outcome. Part B trials from the 1990s and later used a single large dose of thrombolytic drug (80 to 100 mg rt-PA), given intravenously (IV) or intra-arterially (IA) within three, six, nine, or 24 hours of stroke. The primary end point in Part B of the study was a favorable outcome, defined as complete or nearly complete neurological recovery 3 months after stroke. Favorable outcomes were achieved in 31% to 50% of patients treated with rt-PA, as compared to 20% to 38% of patients given placebo (Kwiatkowski TJ, et al., 1999). The benefit was similar one year after stoke. The major risk of treatment was symptomatic intracranial hemorrhage which occurred in 6.4% of patients treated with rt-PA and 0.6% of patients given placebo (Marler JR, et al., 2007).

In 2008, European Cooperative Acute Stoke Study (ECASS)-3, a multi-center, prospective, randomized, placebo-controlled trial, studied the administration of rt-PA between three and 4.5 hours of stroke symptom onset (Hacke W, et al., 2008). The trial enrolled 418 patients treated with rt-PA per the current dosing guidelines (i.e., 0.9 mg/kg (maximum of 90 mg) with 10% given as an initial IV bolus and the remainder infused over one hour) and compared them with 403 who were given placebo. The frequency of the primary efficacy outcome (i.e., modified Rankin Scale score of 0 to 1 at 90 days after treatment) was significantly greater with rt-PA (52.4%) than with placebo (45.2%; OR 1.34, 95% CI 1.02 to 1.76; risk ratio 1.16, 95% CI 1.01 to 1.34; P= 0.04). The point estimate for the degree of benefit seen in ECASS-3 (OR for global favorable outcome, 1.28, 95% CI 1.00 to 1.65) was less than the point estimate of benefit found in the pool of patients enrolled for 0 to 3 hours after stroke symptom onset in the NINDS study (OR 1.9, 95% CI 1.2 to 2.9).

Currently, researchers, along with the field, continue to debate the risk of intracerebral hemorrhage with IV rt-PA in certain patient populations; however, the benefit of improved functional outcomes, and potential improvements in quality of life outweighs the decision to with hold treatment of the ischemia (Saposnick, 2012).

Although the expanded timeframe of 3 to 4.5 hours for thrombolytic therapy was found to be effective for ischemic stroke patients and without significant increase in hemorrhagic events, a treatment target of 3 hours remains the accepted recommendation as studies have shown the potential opportunity for improved outcomes is greater with earlier treatment. Delays in evaluation and initiation of therapy for eligible patients with acute ischemic stroke should be avoided.

1c.4. Citations for data demonstrating high priority provided in 1a.3

•Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A

Guideline From the American Heart Association/American Stroke, Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1655-1711.
• Adams H, Adams R, Del Zoppo G, Goldstein LB. American Heart Association/American Stroke Association Guidelines Update A Scientific Statement From the Stroke Council of the Guidelines for the Early Management of Patients With Ischemic Stroke: 2005, Stroke.2005;36;916-923.
• Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• del Zoppo GJ, Saver JL, Jauch EC, Adams HP. Expansion of the Time Window for Treatment of Acute Ischemic Stroke With Intravenous Tissue Plasminogen Activator: A Science Advisory From the American Heart Association/ American Stroke Association. Stroke. 2009;40:2945-2948.
• George M, Xin T, McGruder H, Yoon P, Rosamond W, Winquist A, Hinchey J, Wall H, Pandey D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Gidetti D, et. al. Thrombolysis with Alteplase 3 to 4.5 hours after acute ischemic stroke. The European Cooperative Acute Stroke Study (ECASS) Investigators. NEJM. 2008;359(13):1317-29.
• Jauch EC, Saver JL, Adams HP, Bruno A, Connors JJ, Demaerschalk BM, Khatri P, McMullan PW, Qureshi AI, Rosenfield K, Scott PA, summers DR, Wang DZ, Wintermark M, Yonas, H. Guidelines for the early management of patients with acute ischemic stroke: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44:893-899.
• Kwiatkowski TG, Libman RB, Frankel M, Tilley BC, Morgenstern LB, Lu M, Broderick JP, Lewandowski CA, Marler HR, Levine SR, Brott T. National Institute of Neurological Disorders and Stroke Recombinant Tissue Plasminogen Activator Study Group. Effects of tissue plasminogen activator for acute ischemic stroke at one year. NEJM. 1999;340:1781-87.
• Marler JR, Tilley BC, Lu M, Brott TG, Lyden PC, Grotta JC, Broderick JP, Levine SR, Frankel MP, Horowitz SH, Haley EC, Lewandowski CA, Kwiatkowski TP. Early stroke treatment associated with better outcome The NINDS rt-PA Stroke Study. Neurology 2000;55:1649-1655.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e-78-82.
• Saposnik G, Fang J, Kapral MK, Tu JV, Mamdani M, Austin P, Johnston C. The iScore Predicts Effectiveness of Thrombolytic Therapy for Acute Ischemic Stroke. Stroke. 2012;43:1315-1322.
• The ATLANTIS, ECASS, and NINDS rt-PA Study Group Investigators. Association of Outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke Trials. Lancet 2004;363:768-774.
• The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. New England Journal of Medicine 1995;333:1581-1587.
• US Burden of Disease Collaborators. The state of US health, 1990–2010: burden of diseases, injuries, and risk factors. JAMA. 2013;310:591–608

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Health and Functional Status : Functional Status, Prevention, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Appendix_A.1-635876964272987900.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
STK 04: Thrombolytic Therapy has been revised since the last endorsement date. A paragraph was added to the measure rationale referencing the European Cooperative Acute Stroke Study (ECASS) III trial and the safe administration of IV thrombolytic therapy within the extended timeframe of 3 to 4.5 hours after stroke for select patients, as well as, the importance of maintaining the administration target of 3 hours. A new data element, Reason for Extending the Initiation of IV Thrombolytic, was added and the measure algorithm adjusted to exclude patients who receive IV thrombolytic therapy within 3 to 4.5 hours after the time last known well when a patient or medical reason is documented in the medical record. Notes for abstraction previously used to exclude these patients were removed from the data element IV Thrombolytic Initiation. Abstraction guidelines for several data elements were also modified to address questions received from data abstractors. The timeframe for documentation of a Reason for Not Initiating IV Thrombolytic was specified as the day of or day after hospital arrival to prevent late entries and addendums after discharge. The same timeframe was specified for the new data element Reason for Extending the Initiation of IV Thrombolytic. Stand-alone reasons were specified for both reason data elements, and a new abstraction bullet added that system reasons are not acceptable. For the data elements Date Last Known Well, Last Known Well, and Time Last Known Well, abstraction guidelines were clarified to address questions about multiple times or conflicting time documentation in the medical record. A sample of vendors and their hospital customers were surveyed to obtain feedback on these changes prior to implementation.

All ICD-9-CM diagnosis codes and ICD-9-CM procedure codes were converted to ICD-10-CM diagnosis and ICD-10-PCS procedure codes throughout the measure specifications. The Department of Health and Human Services (HHS) mandated that all entities covered by the Health Insurance Portability and Accountability Act (HIPAA) must all transition to a new set of codes for electronic health care transactions on October 1, 2015.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Acute ischemic stroke patients for whom IV thrombolytic therapy was initiated at this hospital within 3 hours (less than or equal to 180 minutes) of time last known well.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Five data elements are used to calculate the numerator:

- Date Last Known Well – The month, date, and year prior to hospital arrival at which the patient was last known to be without the signs and symptoms of the current stroke or at his or her baseline state of health.
- Time Last Known Well – The time (military time) prior to hospital arrival at which the patient was last known to be without the signs and symptoms of the current stroke or at his or her baseline state of health.
- IV Thrombolytic Initiation – Documentation that intravenous (IV) thrombolytic therapy (t-PA) was initiated at this hospital. Allowable values: Yes, No/UTD or unable to determine from medical record documentation.
- IV Thrombolytic Initiation Date – The month, date, and year the IV thrombolytic therapy was initiated to a patient with ischemic stroke at this hospital.
- IV Thrombolytic Initiation Time - The time (military time) for which IV thrombolytic therapy was initiated to a patient with ischemic stroke at this hospital.

Patients are eligible for the numerator population when the IV Thrombolytic Initiation Date and IV Thrombolytic Initiation Time minus Date Last Known Well and Time Last Known Well >/= 0 minutes and </= 180 minutes.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Acute ischemic stroke patients whose time of arrival is within 2 hours (less than or equal to 120 minutes) of time last known well.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Fourteen data elements are used to calculate the denominator:

1. Admission Date – The month, day and year of admission to acute inpatient care.
2. Arrival Date – The earliest documented month, day, and year, the patient arrived at the hospital.
3. Arrival Time - The earliest documented time (military time) the patient arrived at the hospital.
4. Birthdate - The month, day and year the patient was born.
5. Clinical Trial - Documentation that during this hospital stay the patient was enrolled in a clinical trial in which patients with stroke were being studied. Allowable values: Yes or No/UTD.
6. Date Last Known Well – The month, date, and year prior to hospital arrival at which the patient was last known to be without the signs and symptoms of the current stroke or at his or her baseline state of health.
7. Discharge Date – The month day and year the patient was discharged from acute care, left against medical advice or expired during the stay.
8. ED Patient – Documentation that the patient received care in a dedicated emergency department of the facility.
Allowable values: Yes or No/UTD.
9. Elective Carotid Intervention – Documentation demonstrates that the current admission is solely for the performance of an elective carotid intervention (e.g., elective carotid endarterectomy, angioplasty, carotid stenting).
Allowable values: Yes or No/UTD.
10. ICD-10-CM Principal Diagnosis Code - The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.
11. Last Known Well – Documentation of the date and time prior to hospital arrival at which it was witnessed or reported that the patient was last known to be without the signs or symptoms of the current stroke or at his or her baseline state of health.

Allowable values: Yes or No/UTD.
12. Reason for Extending the Initiation of IV Thrombolytic – Physician/APN/PA or pharmacist documentation of a reason for extending the initiation of IV thrombolytic.
Allowable values: Yes or No/UTD.
13. Reason For Not Initiating IV Thrombolytic – Physician/APN/PA or pharmacist documentation of a reason for not initiating IV thrombolytic.
Allowable values: Yes or No/UTD.
14. Time Last Known Well – The time (military time) prior to hospital arrival at which the patient was last known to be without the signs and symptoms of the current stroke or at his or her baseline state of health.

Population: Discharges with ICD-10-CM Principal Diagnosis Code for ischemic stroke as defined in Appendix A, Table 8.1.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
• Less than 18 years of age
• Length of Stay > 120 days
• Enrolled in clinical trials related to stroke
• Admitted for elective carotid intervention
• Time last known well to arrival in the emergency department greater than 2 hours
• Documented reason for extending the initiation of IV thrombolytic
• Documented reason for not initiating IV thrombolytic

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
• The patient age in years is equal to the Discharge Date minus the Birthdate. Patients less than 18 years are excluded.
• The Length of Stay (LOS) in days is equal to the Discharge Date minus the Admission Date. If the LOS is greater than 120 days, the patient is excluded.
• Patients are excluded if "Yes" is selected for Clinical Trial.
• Patients are excluded with ICD-10-PCS procedure codes for carotid intervention procedures as identified in Appendix A, Table 8.3, if medical record documentation states that the patient was admitted for the elective performance of this procedure.
• Patients with time last known well to arrival in the emergency department greater than 2 hours are excluded.
• Patients are excluded if “Yes” is selected for Reason for Extending the Initiation of IV Thrombolytic.
• Patients are excluded if "Yes" is selected for Reason For Not Initiating IV Thrombolytic.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable, the measure is not stratified.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable

S.16. **Type of score:**

Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
1. Start processing. Run cases that are included in the Stroke (STK) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check ICD-10-CM Principal Diagnosis Code
a. If the ICD-10-CM Principal Diagnosis Code is not on Table 8.1, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
b. If the ICD-10-CM Principal Diagnosis Code is on Table 8.1, continue processing and proceed to ED Patient.

3. Check ED Patient
a. If ED Patient is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If ED Patient equals No, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If ED Patient equals Yes, continue processing and proceed to Clinical Trial.

4. Check Clinical Trial
a. If Clinical Trial is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Clinical Trial equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Clinical Trial equals No, continue processing and proceed to Elective Carotid Intervention.

5. Check admitted for Elective Carotid Intervention
a. If Elective Carotid Intervention is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Elective Carotid Intervention equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Elective Carotid Intervention equals No, continue processing and proceed to Arrival Date.

6. Check Arrival Date
a. If the Arrival Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If the Arrival Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
c. If the Arrival Date equals a Non-Unable To Determine (non-UTD) Value, continue processing and proceed to Arrival Time.

7. Check Arrival Time only if the Arrival Date is a Non Unable to Determine (non-UTD) Value
a. If the Arrival Time is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If the Arrival Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
c. If the Arrival Time equals a Non-Unable To Determine (non-UTD) Value, continue processing and proceed to Last Known Well.

8. Check Last Known Well
a. If Last Known Well is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Last Known Well equals No, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Last Known Well equals Yes, continue processing and proceed to Date Last Known Well.

9. Check Date Last Known Well
a. If the Date Last Known Well is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop

processing.
b. If the Date Last Known Well equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
c. If the Date Last Known Well equals a Non-Unable To Determine (non-UTD) Value, continue processing and proceed to Time Last Known Well.

10. Check Time Last Known Well only if the Date Last Known Well is a Non Unable to Determine (non-UTD) Value
a. If the Time Last Known Well is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If the Time Last Known Well equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
c. If the Time Last Known Well equals a Non Unable To Determine (non-UTD) Value, continue processing and proceed to the Timing I calculation.

11. Calculate Timing I only if the Time Last Known Well is a Non Unable to Determine (non-UTD) Value. Timing I, in minutes, is equal to the Arrival Date and the Arrival Time minus the Date Last Known Well and the Time Last Known Well. Calculate Timing I for each case that has a Non Unable to Determine (non-UTD) date and time combination.
a. If the time in minutes is greater than 120, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
b. If the time in minutes is greater than or equal to zero and less than or equal to 120, continue processing and proceed to IV Thrombolytic Initiation.

12. Check IV Thrombolytic Initiation
a. If IV Thrombolytic Initiation is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If IV Thrombolytic Initiation equals No, continue processing and proceed to Reason for Not Initiating IV Thrombolytic.
c. If IV Thrombolytic Initiation equals Yes, continue processing and check IV Thrombolytic Initiation Date.

13. Check Reason for Not Initiating IV Thrombolytic
a. If Reason for Not Initiating IV Thrombolytic is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Reason for Not Initiating IV Thrombolytic equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If Reason for Not Initiating IV Thrombolytic equals No, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.

14. Check IV Thrombolytic Initiation Date
a. If the IV Thrombolytic Initiation Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If the IV Thrombolytic Initiation Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
c. If the IV Thrombolytic Initiation Date equals a Non Unable To Determine (non-UTD) Value, continue processing and proceed to IV Thrombolytic Initiation Time.

15. Check IV Thrombolytic Initiation Time only if the IV Thrombolytic Initiation Date is a Non Unable to Determine (non-UTD) Value
a. If the IV Thrombolytic Initiation Time is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If the IV Thrombolytic Initiation Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
c. If the IV Thrombolytic Initiation Time equals a Non Unable To Determine (non-UTD) Value, continue processing and proceed to the Timing II calculation.

16. Calculate Timing II. Timing II, in minutes, is equal to the IV Thrombolytic Initiation Date and the IV Thrombolytic Initiation Time minus the Date Last Known Well and the Time Last Known Well. a. If the time in minutes is greater than 270, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
b. If the time in minutes is greater than or equal to zero and less than or equal to 270, continue processing and proceed to recheck Timing II.

c. If the time in minutes is less than zero, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.

17. Recheck Timing II a. If the time in minutes is greater than or equal to zero and less than or equal to 180, the case will proceed to a Measure category Assignment of E and will be in the Numerator Population. Stop processing.
b. If the time in minutes is greater than 180 and less than or equal to 270, continue processing and proceed to Reason for Extending the Initiation of IV Thrombolytic.

18. Check Reason for Extending the Initiation of IV Thrombolytic a. If Reason for Extending the Initiation of IV Thrombolytic is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Reason for Extending the Initiation of IV Thrombolytic equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Reason for Extending the Initiation of IV Thrombolytic equals No, the case will proceed to a Measure Category Assignment D and will be in the Measure Population. Stop processing.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter for the measure set cannot sample.

Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

Quarterly Sampling
The Quarterly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing quarterly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If "N" >/= 900, then "n" 180
If "N" 226-899, then "n" 20% of Initial Patient Population size
If "N" 45-225, then "n" 45
If "N" 6-44, No sampling; 100% Initial Patient Population required
If "N" 0-5, Submission of patient level data is not required; if submission occurs, 100% Initial Patient Population required

Monthly Sampling
The Monthly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing monthly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If "N >/= 300, then "n" 60
If "N 76-299, then "n" 20% of Initial Patient Population size
If "N 15-75, then "n" 15
If "N < 15, No sampling; 100% Initial Patient Population required

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the

data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Each data element in the data dictionary includes suggested data sources. The data are collected using contracted Performance Measurement Systems (vendors) that develop data collection tools based on the measure specifications. The tools are verified and tested by Joint Commission staff to confirm the accuracy and conformance of the data collection tool with the measure specifications. The vendor may not offer the measure set to hospitals until verification has been passed.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0437_MeasureTesting_MSF5.0_Data-635905384294673612.doc

NQF #: 0437 NQF Project: Neurology Project

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on this measure are as follows:
170 health care organizations representing various hospital demographics:
17 For Profit; 144 Not for Profit; 9 Government
62 >=300 beds; 71 100-299 beds; 37 <100 beds
142 Urban; 28 Rural
22 Teaching; 148 Non-teaching
States represented in this data collection effort include: AL, AR, AZ, CA, CT, FL, GA, IA, ID, IL, IN, KY, LA, MA, MD, MI, MN, MO, MS, NC, NE, NJ, NM, NY, OH, PA, PR, SC, SD, TN, TX, VA, WA, WI
15 performance measurement systems are used for data transmission to The Joint Commission.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with monthly data transmission to The Joint Commission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of approximately 900 patient records identified at 30 pilot sites over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rates of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify of potential measure enhancements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures.) The following is a list of the major tests conducted on the submitted data for this measure:

- Transmission of complete data
- Usage of data received (to understand if the healthcare organization provides the relevant service to treat the relevant population)
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity

- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors is as follows:

- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Data element agreement rates reported to The Joint Commission for the time period of one year (4Q2010 – 3Q2011) have shown an overall agreement rate of 98.1%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-4.

Data Elements	total 'n' numerator	total 'n' denominator	agreement rate
Clinical Trial	712	714	99.7%
Date Last Known Well	181	206	87.9%
ED Patient	693	695	99.7%
Elective Carotid Intervention	711	714	99.6%
IV Thrombolytic Initiation	215	237	90.7%
IV Thrombolytic Initiation Date	25	27	92.6%
IV Thrombolytic Initiation Time	23	27	85.2%
Last Known Well	417	478	87.2%
Reason for Not Initiating IV Thrombolytic Therapy	155	202	76.7%
Time Last Known Well	157	198	80.0%

These agreement rates are considered to be well within acceptable levels.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* are consistent with the evidence cited in support of the measure focus (*criterion 1c)* and identify any differences from the evidence:
This measure focuses on determining the proportion of ischemic stroke patients who arrive at this hospital within two hours of time last known well and for whom IV t-PA was initiated at this hospital within three hours of time last known well. This measure is consistent with the AHA guideline recommendations for thrombolytic threrapy administration within three hours of onset of ischemic stroke.

IV t-PA (i.e., Activase, Alteplase, recombinant tissue plasminogen activator) is the only approved IV thrombolytic for stroke as specified in the data element definition for "IV Thrombolytic Initiation". Patients less than 18 years of age that have a length of stay (LOS) more than 120 days, or enrolled in a clinical trial related to stroke are excluded. In addition, patients who were discharged to a health care facility for hospice care, home for hospice care, who expired or who left against medical advice are excluded to harmonize with other CMS/Joint Commission measures.

There are three exclusions that do not appear in the guideline recommendations. First, patients with a planned admission to the hospital for an elective carotid intervention which restores blood flow to the brain and prevents stroke, such as carotid endarterectomy or carotid stenting, are excluded from the measure. Patients whose arrival time in the emergency room is greater than two hours from time last known well are also excluded because operationally it is unlikely that IV t-PA can be administered within 3 hours. Lastly, patients with a contraindication to thrombolytic therapy or a documented reason why therapy was not initiated are excluded from the measure.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:
At the time this measure was originally tested measure validity was assessed via survey and focus groups of hospitals participating in

the pilot test. All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant.

Since the measure has been in national use, continued validity of the measure has been determined through analysis of feedback from measure users. The Joint Commission provides a web-based application with which measure users can provide feedback regarding appropriateness of measure specifications, request clarification of specifications, and/or provide other comments pertinent to the measure. This feedback is systematically, continually, reviewed in order to identify trends and to identify areas of the measure specifications that require clarification or revision. Additionally, Joint Commission staff continually monitors the national literature and environment in order to assess continued validity of this measure. An expert technical advisory panel (TAP) meets on a quarterly basis in order to assess continued validity of the measure. And finally, the conversion from ICD-9-CM diagnosis codes to ICD-10-CM diagnosis codes has been completed and reviewed by the Technical Advisory Panel for validity. The panel has determined that the intent of the measure has not changed as a result of the conversion. The crosswalk will also be posted in a future version of the specifications manual for public comment, and results of feedback will be reviewed and incorporated into the measure specifications where indicated.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:
Analysis of feedback obtained via our automated feedback system reveals 121 submissions regarding specifications for this measure over the past year. Predominantly, questions involve the data elements Last Known Well, Date Last Known Well, and Time Last Known Well. Non-specific time documentation, such as time ranges (e.g., 1 – 2 hours) and references to a time of day, (e.g., morning, last night), as well as multiple times documented in the medical record by the same or different individuals, challenge abstractors. Stroke severity and fluctuation in presenting symptoms raise questions as to when the clock starts and when it stops. Users have indicated that the hierarchy for date and time documentation provided in the data element notes for abstraction is helpful in such situations. Additionally, abstraction guidelines were added with examples to clarify that time last known well must be prior to arrival time at the hospital because the measure rate calculation cannot accommodate a negative number.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:
Measure exclusions that were not derived directly from the evidence are presented below. Please note that these are population exclusions that are necessary to ensure consistency in all measures in this 8 measure set.

These exclusions were analyzed for frequency of occurrence. An issue that is of great concern to users of this measure is that due to the presence of exceptions to the measure, attainment of a 100% measure rate is not possible. Because of the role of this measure in the current Joint Commission accreditation process this is especially troubling to measure users. This concern is the basis for a number of the non-evidence-based exclusions to these measures. The following measure exclusions that were not derived directly from the evidence are as follows:

1. Patients < 18 years old
2. Patients who have a length of stay (LOS) greater than 120 days
3. Patients enrolled in clinical trials
4. Patients admitted for Elective Carotid Intervention
5. Patients with a documented Reason For Not Initiating IV Thrombolytic

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
N=39,812

1. Patients < 18 years old = 0%
2. Patients who have a length of stay (LOS) greater than 120 days = 0%
3. Patients enrolled in clinical trials =0.39%

4. Patients admitted for Elective Carotid Intervention = 9.96%
5. Patients with a documented Reason For Not Initiating IV Thrombolytic = 0.95%

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Not Applicable

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:
Not Applicable

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:
Not Applicable

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:
Not Applicable

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:
The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of an HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
STK-4 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10th Percentile
3Q2011	104	0.64221	0.40419	1	1	0.8	0.25	0
2Q2011	114	0.65061	0.40260	1	1	0.85714	0.33333	0
Q12011	116	0.6109	0.42108	1	1	0.775	0.0625	0
4Q2010	105	0.65572	0.40568	1	1	0.84615	0.33333	0
3Q2010	104	0.60829	0.42134	1	1	0.75	0	0
2Q2010	101	0.61933	0.38988	1	1	0.66667	0.33333	0
1Q2010	76	0.53939	0.41284	1	1	0.66667	0	0
4Q2009	39	0.54417	0.41298	1	1	0.625	0.03704	0

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)
2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*: Multiple data sources are not used for this measure. 2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*: Not applicable 2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*: Not applicable
2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)
2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: The measure is not stratified for disparities. 2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain: Although some evidence exists that minorities are less likely to receive thrombolytic therapy, there are no plans to stratify the measure. The Joint Commission does not currently capture gender-specific data, or data elements for race or ethnicity because these data elements have not been shown to be reliably collectable due to the fact that no national standardized definitions exist for these data elements. Also, not all hospitals collect race and ethnicity. In the future, it may be feasible for The Joint Commission to explore how race and ethnicity and other relevant disparity data, might be collected reliably in the future.
2.1-2.3 Supplemental Testing Methodology Information:
Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met? (*Reliability and Validity must be rated moderate or high*) Yes☐ No☐ Provide rationale based on specific subcriteria:
If the Committee votes No, STOP

NEW (since 2012 endorsement): Data for Empirical Testing

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

- ☐ **Empirical validity testing**
- ☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The data used to measure validity consists of one year of data (third, and fourth quarter 2014 and first and second quarter for 2015) extracted from The Joint Commission warehouse. The hospital selection was based on those hospitals that reported 12 months of data and had 30 or more denominator cases for the year.

Measure convergent and criterion validity was assessed using hospitals patient level data from The Joint commission warehouse. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous face validity testing. The patient population was comprised of all patients discharged with an ICD-9-CM Principal Diagnosis Code for as defined in Appendix A, Table 8.1 or Table 8.2.

Conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications. The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015.

We conducted a secondary analysis to help interpret results; correlation with other stroke process measures.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Descriptive statistics for the STK measures:
N=1,318 hospitals
n= 2, 206,379 patients records

STK-4
Median: 93%
Percentile 10%: 0%
Percentile 25%: 75%
Percentile 75%: 100%
Percentile 90%: 100%

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations							
	STK_1	STK_2	STK_3	STK_4	STK_5	STK_6	STK_10
STK_1	1.00000 1318	0.55423 <.0001 1318	0.30311 <.0001 1302	0.37009 <.0001 1144	0.52785 <.0001 1318	0.53082 <.0001 1318	0.45291 <.0001 1318
STK_2	0.55423 <.0001 1318	1.00000 1318	0.37566 <.0001 1302	0.28169 <.0001 1144	0.53355 <.0001 1318	0.47326 <.0001 1318	0.29578 <.0001 1318
STK_3	0.30311 <.0001 1302	0.37566 <.0001 1302	1.00000 1302	0.25551 <.0001 1137	0.25665 <.0001 1302	0.28198 <.0001 1302	0.15819 <.0001 1302
STK_4	0.37009 <.0001 1144	0.28169 <.0001 1144	0.25551 <.0001 1137	1.00000 1144	0.22516 <.0001 1144	0.43717 <.0001 1144	0.41076 <.0001 1144
STK_5	0.52785 <.0001 1318	0.53355 <.0001 1318	0.25665 <.0001 1302	0.22516 <.0001 1144	1.00000 1318	0.35302 <.0001 1318	0.39079 <.0001 1318
STK_6	0.53082 <.0001 1318	0.47326 <.0001 1318	0.28198 <.0001 1302	0.43717 <.0001 1144	0.35302 <.0001 1318	1.00000 1318	0.45420 <.0001 1318
STK_10	0.45291 <.0001 1318	0.29578 <.0001 1318	0.15819 <.0001 1302	0.41076 <.0001 1144	0.39079 <.0001 1318	0.45420 <.0001 1318	1.00000 1318

The correlations of the stroke measures are all positively correlated and statistically significant indicating that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)
Overall, the positive inter-correlations indicate convergent validity of the measures.
STK-04 is positively correlated with other stroke evidence-based processes of care.

2b3. EXCLUSIONS ANALYSIS .
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patient who have a Length of Stay greater than 120 days

- Patient enrolled in a Clinical Trial
- Patients admitted for *Elective Carotid Intervention*
- *Time Last Known Well* to arrival in the emergency department greater than 2 hours
- Patients with a documented *Reason For Extending the Initiation of IV Thrombolytic*
- Patients with a documented *Reason For Not Initiating IV Thrombolytic*

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are given below.

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%
10th Percentile: 0.24%
Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for *Elective Carotid Intervention*
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.5%
90th Percentile: 25.3%
Maximum: 95.2%

Exclusion: *Time Last Known Well* to arrival in the emergency department greater than 2 hours
None

Exclusion: Patients with a documented *Reason For Not Initiating IV Thrombolytic*
Overall Occurrence n = 73,966
Overall Occurrence Percentage: 24%
Minimum: 0.42%
10th Percentile: 5.5%
Median: 14.6%
90th Percentile: 54%
Maximum: 92%

Exclusion: Patients with a documented *Reason For Extending the Initiation of IV Thrombolytic*
Overall Occurrence = 2,208
Overall Occurrence Percentage: 0.70%
Minimum: 0.09%
10th Percentile: 0.25%
Median: 0.69%
90th Percentile: 2.3%
Maximum: 35.2%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data*

collection and analysis. *Note*: ***If patient preference is an exclusion****, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The median frequency of exclusions range from low to moderate. The distribution of exclusions across hospitals is generally narrow indicating that the occurrence is random and likely would not bias performance results, although the percentage of patients excluded may differ depending on whether the hospital was a stroke center.

For criterion validity, it is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a Length of Stay greater than 120 days
Rationale: Included to harmonize with other CMS/Joint Commission aligned measures.

Patients enrolled in a Clinical Trial
Rationale: Thrombolytic therapy as specified by this measure may compromise a clinical trial.

Patients admitted for *Elective Carotid Intervention*
Rationale: Patients who are not admitted for treatment of an acute stroke may not require thrombolytic therapy

Patients with a documented Reason For Not Initiating IV Thrombolytic
Rationale: It is inappropriate to treat patients who have a documented reason or contraindication to thrombolytic therapy

Patients with a documented Reason For Extending the Initiation of IV Thrombolytic
Rationale: It may be appropriate to treat select patients within 3 to 4.5 hours

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other

If other: Data element allowable values are selected, either manually or electronically, from clinical and coded data available in medical record documentation. All medical record documentation is used in the abstraction process. Vendor data collection tools are used to import data elements needed for measure rate calculation.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract the electronic version of this measure, so continues to offer this chart-abstracted version which allows for data capture from unstructured data fields. All data elements needed to compute the STK-4 performance measure score have been retooled for capture from electronic sources. Annual updates are performed to match the eCQM specifications to the current version of the chart-abstracted specifications.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

At the present time, hospitals using this performance measure generally collect measure data via manual review of the paper medical record, the EHR or a combination of both. Collected data are submitted to The Joint Commission on a quarterly basis, by way of contracted performance measurement system vendors, as described previously. Specifications for this measure are freely available to anyone who wishes to use the measure. Feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As described above, as feedback from measure users has indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

There are no fees or licensing requirements to use The Joint Commission performance measures, all of which are in the public

domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Quality Check® http://www.qualitycheck.org/consumer/searchQCR.aspx Hospital Compare https://www.medicare.gov/hospitalcompare/search.html Public Health/Disease Surveillance Paul Coverdell National Acute Stroke Registry http://www.cdc.gov/dhdsp/programs/stroke_registry.htm Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program http://www.jointcommission.org/ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Annual Report – Improving America's Hospitals http://www.jointcommission.org/annualreport.aspx Quality Improvement (Internal to the specific organization) Disease-Specific Care Certification for Comprehensive Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx http://www.jointcommission.org/certification/dsc_home.aspx Disease-Specific Care Certification for Primary Stroke Centers

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

• Name of program and sponsor: Quality Check®; The Joint Commission

• Purpose: A public website that allows consumers to: search for accredited and certified organizations by city and state, by name or

by zip code (up to 250 miles); find organizations by type of service provided within a geographic area; download free hospital performance measure results; and, print a list of Joint Commission certified disease-specific care programs and health care staffing firms.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

• Name of program and sponsor: Hospital Compare; Centers for Medicare & Medicaid Services
• Purpose: A public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4000+ Medicare-certified hospitals (2015)

• Name of program and sponsor: Paul Coverdell National Acute Stroke Registry; Centers for Disease Control and Prevention
• Purpose: A national registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.
• Geographic area and number and percentage of accountable entities and patients included: 11 states; 403 hospitals (CDC, 2014)

• Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
• Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 Medicare-certified hospitals (2015)

• Name of program and sponsor: Annual Report-Improving America's Hospitals; The Joint Commission
• Purpose: The annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

• Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

• Name of program and sponsor Disease-Specific Care Certification for Primary Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
• Geographic area and number and percentage of accountable entities and patients included Nationwide; 1079 hospitals

• Name of program and sponsor Hospital Accreditation Program; The Joint Commission
• Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
• Geographic area and number and percentage of accountable entities and patients included Nationwide; 3300 Joint Commission-accredited hospitals (2014)

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for*

implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.)
Not applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

• Progress (trends in performance results, number and percentage of people receiving high-quality healthcare) The rate of intravenous thrombolytic (t-PA) administration has improved by 57.8% over the past five years based on Joint Commission ORYX performance measurement data. This trend is consistent with increases in the rate of IV t-PA administration reported in multiple published studies. Other trends identified in the literature include: an increasing interest in the concept of treating patients with milder symptoms or rapid improvement; an increased proportion of very old (> 85 years) acute ischemic stroke patients receiving tPA; an increased proportion of young acute ischemic stroke patients (mean age 37 years), in particular young blacks, receiving tPA; and, a slight increase in the proportion of women receiving tPA. Additionally, an increased trend in the rate of tPA utilization for urban patients treated at urban hospitals, as well as, rural patients treated at rural hospitals has been reported.

• Geographic area and number and percentage of accountable entities and patients included Nationwide; 1099 hospitals; 14,907 patients (The Joint Commission, 2014)

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Guidelines for the expansion of the time window for treatment of acute ischemic stroke with intravenous tissue plasminogen activator (American Heart Association/American Stroke Association, 2009) have been challenging for STK-4 Thrombolytic Therapy which includes only those patients who receive IV t-PA within 3 hours in the numerator population. The updated guideline recommendation for IV t-PA administration within 3 to 4.5 hours after stroke generated discussion about the STK-4 measure construct and the possibility of revising the STK-4 numerator to include patients treated within 4.5 hours. This recommendation also generated concern that thrombolytic therapy initiation would be delayed for patients eligible to receive treatment within 3 hours and result in poorer outcomes for stroke patients. Since earlier initiation is associated with better neurological outcomes, the STK-4 numerator has remained unchanged; however, the measure specifications have been modified to clearly exclude patients who are treated with IV t-PA within 3 to 4.5 hours and have a documented medical or patient reason for the delay in initiation. An unexpected benefit of this change is that it promotes healthcare staff awareness that earlier initiation of thrombolytic therapy is better for stroke patients and also preserves the goal of the NINDS investigators to initiate thrombolytic treatment for eligible patients within 3 hours.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually

both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0164 : Fibrinolytic Therapy received within 30 minutes of hospital arrival
0242 : Stroke and Stroke Rehabilitation: Tissue Plasminogen Activator (t-PA) Considered
0288 : Fibrinolytic Therapy Received Within 30 Minutes of ED Arrival
1952 : Time to Intravenous Thrombolytic Therapy

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0242 : Stroke and Stroke Rehabilitation: Tissue Plasminogen Activator (t-PA) Considered
American Medical Association - Physician Consortium for Performance Improvement (AMA-PCPI) – no longer NQF endorsed

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Measures 0288 and 0164 are AMI (Acute Myocardial Infarction) measures. They are part of the Centers for Medicare & Medicaid Services/The Joint Commission aligned measures relating to the administration of fibrinolytic therapy for hospital inpatients and are harmonized with 0437 to the extent that the measures utilize some of the same data elements. The target population for 0288 and 0164 is inpatients with an ICD-10-CM Principal Diagnosis Code for acute myocardial infarction. The target population for 0437 differs in that it includes patients hospitalized for acute ischemic stroke. In addition, the evidence around the timeframe for administration of therapy is different for the AMI and ischemic stroke populations, and 0288 and 0164 include administration of lytic drugs other than activase/alteplase/IV t-PA/recombinant tissue plasminogen activator (rt-PA). Measure 0164 will be removed from the CMS/The Joint Commission aligned measures starting with 01/01/2016 discharges. The target population for measure 1952 from the American Heart Association/American Stroke Association also includes patients hospitalized for acute ischemic stroke; however, the measure captures average door-to-needle time and uses a target of less than 60 minutes rather than the proportion of patients who arrive within 2 hours and receive t-PA within 3 hours of time last known well. Measure 0242 is a physician performance measure with a targeted population of ischemic stroke patients identified through CPT codes and could extend to the outpatient setting. This measure evaluates physician practice as opposed to hospital processes. It is no longer NQF-endorsed

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not Applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Ann, Watt, awatt@jointcommission.org, 630-792-5944-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Karen, Kolbusz, kkolbusz@jointcommission.org, 630-792-5931-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel (TAP) is to provide advisory oversight in literature review, measure content and maintenance of the specifications.
Members are:

Harold P. Adams, Jr., MD
University of Iowa Health Care
Iowa City, IA

Mark J. Alberts, MD
University of Texas Southwestern
Dallas, TX

Anne W. Alexandrov, RN
Health Outcomes Institute
Fountain Hills, AZ

Nadine Allyn, RD
American Heart Association
Dallas, Texas

Mary G. George, MD
Centers for Disease Control and Prevention
Atlanta, GA

Martin Gizzi, MD
Legacy Health System
Portland, OR

Marianna Gorbaty
Mathematica Policy Research, Inc.
Washington, DC

Judith Hinchey, MD
Tuft's Medical School
Steward St. Elizabeth's Medical Center
Boston, MA

Edward C. Jauch, MD
Medical University of South Carolina
Charleston, SC

Irene Katzan, MD
Cleveland Clinic
Cleveland, OH

Kimberly E. Levasseur-Franklin, PharmD
Northwestern Memorial Hospital
Chicago, IL

Kathy Morrison, RN
Penn State Hershey Medical Center
Hershey, PA

Marilyn M. Rymer, MD
University of Kansas
Kansas City, MO

Jeffrey Saver, MD
UCLA Medical Center
Los Angeles, CA

Lee H. Schwamm, MD
Harvard Medical School
Massachusetts General Hospital
Boston, MA

Deborah, Summers, RN
St. Luke's Brain and Stroke Institute
Kansas City, MO

Osama O. Zaidat, MD
St. Vincent Mercy Hospital
Toledo, OH

Richard D. Zorowitz, MD
Medstar National Rehabilitation Network
Washington, DC

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2009
Ad.3 **Month and Year of most recent revision:** 10, 2015
Ad.4 **What is your frequency for review/update of this measure?** Biannual
Ad.5 **When is the next scheduled review/update for this measure?** 07, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0438
Measure Title: STK 05: Antithrombotic Therapy By End of Hospital Day Two
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of ischemic stroke patients who had antithrombotic therapy administered by end of hospital day two (with the day of arrival being day 1).
This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-4: Thrombolytic Therapy, STK-6: Discharged on Statin Medication, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs.
Developer Rationale: Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. The effectiveness of early antithrombotic administration in reducing stroke mortality recurrence rates has been demonstrated in several large clinical trials. Data suggest that antithrombotic therapy should be administered by the end of hospital day two to reduce stroke mortality and morbidity as long as no contraindications exist.
Healthcare organizations that track early antithrombotic therapy for internal quality improvement purposes have seen an improvement in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

Numerator Statement: Ischemic stroke patients who had antithrombotic therapy administered by end of hospital day two.
Denominator Statement: Ischemic stroke patients
Denominator Exclusions:
- Less than 18 years of age
- Duration of Stay < 2 days
- Length of Stay > 120 days
- Comfort measures only documented on the day of or day after hospital arrival
- Enrolled in clinical trials related to stroke
- Admitted for elective carotid intervention
- IV OR IA thrombolytic therapy administered at this hospital or within 24 hours prior to arrival
- Documented reason for not administering antithrombotic therapy by end of hospital day 2

Measure Type: Process
Data Source: Electronic Clinical Data, Paper Medical Records
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2008 **Most Recent Endorsement Date:** Nov 01, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Summary of prior review in 2012:

- The body of evidence consistently supports that early antithrombotic therapy reduces the risk of non-fatal stroke and death for patients who have experienced an acute ischemic stroke. There is no evidence refuting the benefit of this recommendation.
- Class I, Level of Evidence A – AHA ASA 2007 Guidelines for the Early Management of Adults with Ischemic Stroke, page 1681.
    - The oral administration of aspirin (initial dose is 325 mg) within 24 to 48 hours after stroke onset is recommended for treatment of most patients (Class I, Level of Evidence A). This recommendation has changed in that a dose of aspirin is now included.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☐ **The developer provided updated evidence for this measure:**

Guidance from the Evidence Algorithm:

Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency high(Box 5) → Box 5a High

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and voting on Evidence?*

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer provides the following national trend data:

	CY 2010	CY2011	CY 2012	CY 2013	CY 2014
# hospitals	137	157	158	263	1,300
# cases (den)	18,728	22,328	22,258	33,651	162,275
National aggregate rate	0.97	0.98	0.98	0.98	0.99
Mean hospital rate	0.96	0.96	0.98	0.98	0.98
50th percentile	0.98	0.99	0.99	0.99	0.99
10th and 90th percentiles	0.91 (10th) 1.00 (90th)	0.92 (10th) 1.00 (90th)	0.95 (10th) 1.00 (90th)	0.96 (10th) 1.00 (90th)	0.96 (10th) 1.0 (90th)

- Participation in this measure has grown significantly in the past five years. National hospital performance seems to be very high with little to no room for improvement.
- In 2012, the Committee noted the high performance for this measure (98%), but agreed that even a relatively small increase in performance would affect a large number of patients.

Disparities:

- The developer does not provide disparities data from use of this measure.
- Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age."

Questions for the Committee:

- *Performance over five years has been consistently high. How much further improvement in performance is likely using this measure?*
- *Does a gap in care still exist that warrants this national performance measure?*
- *Is there evidence that a gap in care still exits for minority populations? How can this measure be used to better understand disparities in care and outcomes for certain groups?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **This is a process measure to assess rate of administration of aspirin by end of hospital day two in patients with ischemic stroke. Evidence is good that aspirin prevents additional strokes if administered early and recommendations from professional groups are strong for administration within 24-48 hours of acute event. Evidence for measure has not changed. Agree with preliminary ratings.

**No new evidence since 2012

**This is a maintenance process measure. I am not aware of any new evidence since the measure was last evaluated.

**No evidence changes reported since last evaluation. No vote

1b. Performance Gap

Comments: **Performance over 5 years has been very high. No evidence is provided that a gap exists in this measure for minority populations.

Agree little opportunity for improvement.

**99% participation with little room for improvement
**Performance data was provided and does NOT demonstrate a significant gap in care. Data on disparities was NOT provided, but if it were available, this measure could be used to determine if there are differences in performance in minority populations.
**Higher participation over time. Only a small room for improvement - though this may represent significant numbers of individuals affected.
No data on disparities. Very possible the disparities may account for a large portion of the room for improvement.

1c. High Priority (previously referred to as High Impact)

Comments: **Clear construction.
**NA

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): medical record abstraction (paper or electronic). This is not an eMeasure.

Specifications:

- One data element defines the numerator and 10 data elements define the denominator.
- Diagnosis and procedure codes have been converted from ICD-9 to ICD-10-CM diagnosis and ICD-10-PCS procedure codes. See attached spreadsheet Appendix A Table 8.1 (ischemic stroke).
- A conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications:
    - "The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015 developer advisory panel determined that the intent of the measure was changed by the conversion to ICD-10 and a crosswalk was posted."
- Sampling monthly or quarterly is allowed; instructions for sampling are provided.
- A web-based data collection tool has been developed but is not described. Hospitals are not required to use this tool.
- The measures is specified at the hospital level of analysis.
- The developer states that there have been no significant changes to this measure since the last endorsement date.
- In 2012, one Committee member requested clarification regarding the definition of the antithrombotic regimen—specifically whether it includes only aspirin therapy. The developer noted that all FDA-approved or guideline-endorsed antithrombotic agents are listed in a data dictionary for the measure.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Inter rater-reliability of the data elements for one year (4Q2010-3Q2011) in 77 hospitals and 739 patient records showed an overall agreement rate of 97.2%.
  - Five data elements found ≥95% agreement: Antithrombotic Therapy Administered by End of Hospital Day 2 (95.0%), Clinical Trial (99.7%), Comfort Measures Only (99.3%), Elective Carotid Intervention (99.6%), and IV OR IA Thrombolytic Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival (97.5%).
  - One data element found < 95% agreement: Reason for Not Administering Antithrombotic Therapy by End of Hospital Day 2 (93.6%).
  - No Kappa scores provided.

- The 2012 Committee expressed no concerns regarding reliability of this measure.

Describe any updates to testing: No new information provided

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing: see above

Results of reliability testing : see above

Guidance from the Reliability Algorithm :

Precise specifications (Box 1) → empirical testing as specified (Box 2) → reliability testing with patient-level data elements (Box 8) → assessed all critical data elements (Box 9) → high/moderate certainty the data elements are reliable (Box 10) → Moderate (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Questions for the Committee:

- *The developer has not provided any new reliability testing. Does the Committee agree there is no need for repeat discussion and voting on Reliability?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Face validity of the data was assessed via survey and focus groups of hospitals participating in the pilot test.
- The developer reports that "All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant."
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.
- Ongoing feedback from measure users that is systematically reviewed to identify trends and/or measure specifications that require clarification or revision. Predominantly, questions regarding clarification of the data element Reason for Not Prescribing Antithrombotic Therapy By End of Hospital Day 2, with respect to both acceptable reasons and the timeframe for documentation. Inquiries about new anticoagulants dabigatran and rivaroxaban pertaining to inclusion of patients taking these medications in the numerator were also received. The medication Table 8.2 has been updated to include these medications.

Describe any updates to validity testing – New empirical validity testing data is presented

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- Pearson Correlation Coefficients were calculated for seven stroke measures for hospitals that reported 12 months of data and had 30 or more denominator cases for the year (2014-2015).Presumably high performing hospitals would do well on all stroke performance measures.

Validity testing results:

- Analysis of 1,318 hospitals and 2,206,379 patients records generated a table of Pearson Correlation Coefficient results that shows a statistically significant (P<.0001), positive correlation of this measure with six other stroke performance measures supporting the hypothesis that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients are excluded from the measure for the following reasons:
  - Patients less than 18 years of age
  - Patients who have a Duration of Stay less than 2 days
  - Patient who have a Length of Stay greater than 120 days
  - Patients with Comfort Measures documented on day of or day after arrival
  - Patient enrolled in a Clinical Trial

- Patients admitted for Elective Carotid Intervention
- Patients discharged prior to the end of hospital day 2
- Patients with IV OR IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival
- Patients with a documented Reason for Not Administering Antithrombotic Therapy by End of Hospital Day 2

- New data on the frequency of exclusions is presented for 2, 206,379 admissions in 1,318 hospitals:
  - Comfort Measures Only: 4.46% (range 0.31-30.8%)
  - Clinical Trial: 0.25% (range 0.08-10.0%)
  - Patients admitted for Elective Carotid Intervention: 11.8% (range 0.16-95.2%)
  - Patients with IV OR IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival: 8.04% (range 0.25-47.4%)
  - Patients with a documented Reason for Not Administering Antithrombotic Therapy by End of Hospital Day 2: 7.11% (range 0.35-76.9%)

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Performance data is presented above under opportunity for improvement. Complete details of the data is presented in 1b.
- The developer explains their approach: "There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations."

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods: Not Applicable

2b7. Missing Data

- The developer describes the handling of missing data in the measure specifications (S.22) but does not provide information on the frequency of missing data or potential impact on measure results.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → potential threats to validity mostly assessed (Box2) →empirical validity testing (Box 3) → measure score testing (Box 6) → sound conceptualization and hypotheses (Box 7) →moderate confidence that score are a valid indicator of quality(Box 8a) → High

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **No concerns.

**No concerns
**Data elements are clearly defined. No concerns.
**Specifications are consistent with the evidence.
**Very high - includes face validity. No need to vote.

2a2. Reliability Testing

Comments: **Agree that score from this measure is an indicator of quality.
**No concerns
**New empirical validity testing provided: Analysis of 1,318 hospitals and 2,206,379 patients records generated a table of Pearson Correlation Coefficient results that shows a statistically significant (P<.0001), positive correlation of this measure with six other stroke performance measures supporting the hypothesis that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.
**Very high - no need for vote.
May warrant examination for different anti-thrombolitics other than aspirin.
**Agree- no need for discussion on reliability.
**No concerns
**Prior reliability testing from 2012 review showed inter rater-reliability of the data elements for one year (4Q2010-3Q2011) in 77 hospitals and 739 patient records with an overall agreement rate of 97.2%. Only one data element had < 95% agreement, Reason for Not Administering Antithrombotic Therapy by End of Hospital Day 2 (93.6%). No new info provided since 2012 review.

2b2. Validity Testing

Comments: **No
**No concerns re-threats to validity
**The developer does not provide info on frequency of missing data or potential impact.
**Exclusions are appropriate
No Risk Adjustments
Missing data - proposal for handling, no data on frequency.
**Agree that score from this measure is an indicator of quality.
**New empirical validity testing provided: Analysis of 1,318 hospitals and 2,206,379 patients records generated a table of Pearson Correlation Coefficient results that shows a statistically significant (P<.0001), positive correlation of this measure with six other stroke performance measures supporting the hypothesis that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Agree- no need for discussion on reliability.
**No concerns
**Prior reliability testing from 2012 review showed inter rater-reliability of the data elements for one year (4Q2010-3Q2011) in 77 hospitals and 739 patient records with an overall agreement rate of 97.2%. Only one data element had < 95% agreement, Reason for Not Administering Antithrombotic Therapy by End of Hospital Day 2 (93.6%). No new info provided since 2012 review.
**Very high - no need for vote.
**No concerns re-threats to validity
**The developer does not provide info on frequency of missing data or potential impact.
**2d. Yes

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is medical record abstraction: not all hospitals are able to generate the data electronically so a paper-based option is available.
- Some data elements are in electronic form.
- Data collection burden is significant.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Actions common in clinical care, charted, should be easy to extract.

**Data elements are manually abstracted from medical records, creating a significant burden.

**Data are in records. Not all hospitals have ability to enter the data in EHR, thus warranted preservation of this chart version.

Chart data warrant additional collection and analysis time. What is the cost-benefit when adherence is so high?

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details :

- The Joint Commission Quality Check®: Public website that allows consumers to search for accredited and certified organizations.
- CDC Paul Coverdell National Acute Stroke Registry: National registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.
- CMS Hospital Inpatient Quality Reporting Program (IQR): The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- The Joint Commission Annual Report-Improving America's Hospitals: Annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care

processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
- The Joint Commission Disease-Specific Care Certification for Comprehensive Stroke Centers: Certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
- The Joint Commission Disease-Specific Care Certification for Primary Stroke Centers: Certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results:
- The developer provided the following progress information on 1,300 hospitals nationwide and 162,275 patients:
  - The rate of early antithrombotic therapy has steadily increased over the past five years based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 98.5%. A modest gap of approximately 4% still exists for the 10th percentile of hospitals.
  - This trend is consistent with increases in the rate of antithrombotic therapy at discharge published by GWTG-Stroke and PCNASR.
  - Significant and important differences between races in rates of early antithrombotic therapy persist for Hispanics and non-Hispanic blacks (Schwamm, 2010, Qian, 3013; CDC 2014).

Unexpected findings (positive or negative) during implementation:
- Omission of early antithrombotic therapy when indicated may result in poorer neurological outcomes for stroke patients. The timeframe specified for antithrombotic administration is the day of or day after hospital arrival. The two-day timeframe is aligned with the guideline recommendation that aspirin should be administered within 24 to 48 hours of stroke onset.
  - A two-day computation is used rather than abstracting time in precise minutes and hours to ease abstraction burden. Antithrombotic therapy must be administered by 2359 on the day after hospital arrival (day two), or a reason for not administering documented; however, this methodology for abstraction may result in a fall-out from the measure if the patient arrives late in the day on day one and antithrombotic therapy is administered within 48 hours of hospital arrival.

Potential harms: None reported

Feedback :
- The developer states that feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As previously noted, measure users have indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

***Questions for the Committee**:*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Simple measure easily implemented and reported.

**Continued use should benefit those pts where disparities still exist

**The measure has been used widely to promote high-quality acute stroke care.

**The data are used in accountability. The already high compliance warrants consideration of further exploration of disparities, relative benefit of other anti-thrombolitics and some of the exclusions.

Criterion 5: Related and Competing Measures

Related or competing measures:

- 0068 : Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet
- 0435 : STK 02: Discharged on Antithrombotic Therapy
- 0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy – American Academy of Neurology – not NQF endorsed

Harmonization:

- Measure 0435, Discharged on Antithrombotic Therapy, is the second (STK-2) measure in The Joint Commission stroke core measure set and also targets the ischemic stroke population; however, the timeframe for antithrombotic administration is different in the two measures. #0435 focuses on the prescription of antithrombotic medications at the time of hospital discharge. All other data elements are aligned between the two measures.
- Measure 0068 is a physician level measure with a different target population - patients with ischemic vascular disease who were discharged alive for acute myocardial infarction (AMI), coronary artery bypass graft (CABG) or percutaneous coronary interventions (PCI).

Pre-meeting public and member comments

No comments were submitted for this measure during the pre-meeting comment period.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0438 NQF Project: Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.

Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The focus of the measure is to decrease stroke recurrence rates for ischemic stroke patients by administering aspirin or another antithrombotic medication within 2 days of symptom onset.
Antithrombotic administered by end of hospital day 2 >> secondary stroke prevention >> decreased morbidity and mortality.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
This measure is consistent with the guidelines recommended by the American Heart Association/American Stroke Association for the early management of adults with ischemic stroke. A small but statistically significant decline in risk of mortality and morbidity has been demonstrated when aspirin is administered within 48 hours after onset of stroke. The primary effects of aspirin are due to reduction of early recurrent stroke rather than limitation of neurological consequences of the stroke. The focus of both the performance measure and the body of evidence supports the need for early administration of antithrombotic therapy.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles*): The two most frequently cited trials in the literature are the International Stroke Trial (i.e., 300 mg of aspirin daily versus control; 1994) and the Chinese Acute Stroke Trial(i.e., 160 mg aspirin daily versus control; 1997). Prior to these trials, there was little to no information about the effects of antiplatelet agents on acute ischemic stroke.

In 2002, the Antithrombotic Trialists' Collaboration published a systematic overview of the relevant literature supporting antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. The collaboration reviewed 287 studies involving 135,000 patients in comparisons of antiplatelet therapy versus control and 77,000 patients in comparisons of different antiplatelet regimens. Relevant trials were identified through electronic database searches (MEDLINE, Embase, Derwent, Scisearch, and Biosis), searching the trial registers of the Cochrane Stroke and Peripheral Vascular Disease Groups; manual searching of journals, abstracts, and proceedings of meetings; scrutinizing the reference lists of trials and review articles; and professional inquiry, including colleagues and representatives of pharmaceutical companies.

The meta-analysis identified seven randomized control trials encompassing 40,821 patients that focused on the effects of antithrombotic therapy in acute ischemic stroke patients. The patients enrolled in these seven trials received on average three weeks of antiplatelet therapy following the onset of acute stroke symptoms. Overall, antiplatelet therapy resulted in a proportional reduction of 11% in vascular events (i.e., non-fatal stroke or death).

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting early antithrombotic administration for patients with acute ischemic stroke is high. Multiple large randomized control studies have been published that studied the effects of aspirin, aspirin in combination with other antithrombotic agents, and alternative antithrombotic agents as monotherapy. As noted by the AHA/ASA and the American Academy of Neurology, the early administration of antithrombotic therapy, the focus of this measure, has been found to protect against stroke recurrence in the initial weeks following ischemic stroke. The administration of aspirin is not a substitute for other acute interventions, especially intravenous administration of thrombolytic therapy (IV t-PA), for the treatment of acute ischemic stroke. The administration of aspirin as an adjunctive therapy within 24 hours of IV t-PA is also not recommended.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of evidence consistently supports that early antithrombotic therapy reduces the risk of non-fatal stroke and death for patients who have experienced an acute ischemic stroke. There is no evidence refuting the benefit of this recommendation.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
According to the Antithrombotic Trialists' Collaboration meta-analysis, there is strong evidence on the immediate hazards and net benefits of early antithrombotic administration for the prevention of stroke recurrence. Four of the seven trials included in the trialists' 2002 meta-analysis separated stroke outcomes into those considered to be due to hemorrhagic events and those of ischemic or unknown origin. Antithrombotic therapy produced an absolute excess of 1.9 hemorrhagic strokes per 1,000 patients, which was counterbalanced by an absolute reduction of 6.9 fewer ischemic strokes per 1,000, yielding an overall risk reduction in the risk of any further stroke of 5.4 per 1,000 patients. The excess risk of major extracranial hemorrhage was estimated at about three excess bleeds per 1,000 patients receiving early antithrombotic therapy. An increased risk of bleeding was associated with concurrent heparin administration (i.e., absolute excess 9 bleeds per 1,000 with heparin versus 2 bleeds per 1,000 without heparin.

As previously stated, the benefits of early antithrombotic therapy outweigh the risks of hemorrhage. Aspirin is a cost-effective therapy (Gaspoz, et al., 2002).

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: American Heart Association. During our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, and provided information that was accessible and met the requirements set out in this measure maintenance form.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.13 Grade Assigned to the Body of Evidence: Class I, Level of Evidence A

1c.14 Summary of Controversy/Contradictory Evidence: The benefit of early antithrombotic therapy post ischemic stroke is undisputed. No position against early antithrombotic therapy was noted in the literature; however, the substitution of early antithrombotic therapy for other acute interventions, (i.e., IV t-PA) is not recommended. Additionally, the administration of antithrombotic medications in conjunction with other antithrombotic agents (e.g., heparin) or soon after IV t-PA administration is inadvisable.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and thrombolytic therapy for ischemic stroke: The Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest. 2004; 126;483S-512S.
- Antithrombotic Trialists' Collaboration. Collaborative meta-analysis of randomized trials of anti-platelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients [published correction appears in BMJ. 2002;324:141]. BMJ. 2002;324:71– 86.
- Brott TG, Clark WM, Grotta JC, et al. Stroke the first hours. Guidelines for acute treatment. Consensus Statement. National Stroke Association. 2000.
- Chen ZM, Sandercock P, on behalf of the AntiThrombotic Trialists Collaboration (ATT). Indications for early aspirin use in acute stroke: a systematic overview of over 40,000 randomised patients. Cerebrovasc Dis. 1998;8(suppl 4):38.
- Chen ZM, Sandercock P, Pan HC, et al. Indications for early aspirin use in acute ischemic stroke: a combined analysis of 40,000 randomized patients from the Chinese acute stroke trial and the international stroke trial. On behalf of the CAST and IST collaborative groups, Stroke 2000;31:1240-1249.
- Coull BM, Williams LS, Goldstein LB, et al. Anticoagulants and Antiplatelet Agents in Acute Ischemic Stroke. Report of the Joint Stroke Guideline Development Committee of the American Academy of Neurology and the American Stroke Association (a Division of the American Heart Association) Stroke. 2002;33:1934 -1942.
- Furie KL, Kasner SE, Adams RJ, Albers GW, et al. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2011;42:246-49.
- Gaspoz J, Coxson PG, Goldman PA, Williams LW, Kuntz KM, Hunink MGM, Goldman L. Cost effectiveness of aspirin, clopidogrel, or both for secondary prevention of coronary heart disease. NEJM. 2002;346:1800-1806.
- Guyatt GH, Akl EA, Crowther M, Gutterman DD, Sch?nemann HJ, and for the American College of Chest Physicians

Antithrombotic Therapy and Prevention of Thrombosis Panel. Executive summary: antithrombotic therapy and prevention of thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest. 2012;141: 34S.
• Johnson ES, Lanes SF, Wentworth CE, Satterfield MH, Abebe BL, Dicker LW. A meta-regression analysis of the dose-response effect of aspirin on stroke. Arch Intern Med. 1999;159:1248 –1253.
• Sandercock P, Collins R, Counsell C, Farrell B, Peto R, Slattery J, Warlow C, International Stroke Trial Collaborative Group: The International Stroke Trial (IST): A randomised trial of aspirin, subcutaneous heparin, both, or neither among 19,435 patients with acute ischaemic stroke. Lancet. 1997;349:1569-1581.
• The ESPS Group. The European Stroke Prevention Study (ESPS): principal end-points. Lancet. 1987:2:1351–1354.

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)*:
AHA/ASA 2007 Guidelines for the Early Management of Adults with Ischemic Stroke, page 1681.
Class I, Level of Evidence A
1. The oral administration of aspirin (initial dose is 325 mg) within 24 to 48 hours after stroke onset is recommended for treatment of most patients (Class I, Level of Evidence A). This recommendation has changed in that a dose of aspirin is now included.

1c.17 Clinical Practice Guideline Citation: Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A Guideline From the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1681.

1c.18 National Guideline Clearinghouse or other URL:
http://www.guidelines.gov/search/search.aspx?term=early+management+of+adults+with+ischemic+stroke

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: This guideline was approved by the American Heart Association Science Advisory and Coordinating Committee on January 6, 2007. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.23 Grade Assigned to the Recommendation: Class I, Level of Evidence A

1c.24 Rationale for Using this Guideline Over Others: The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: High 1c.26 Quality: High1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0438_Evidence_MSF5.0_Data.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. The effectiveness of early antithrombotic administration in reducing stroke mortality recurrence rates has been demonstrated in several large clinical trials. Data suggest that antithrombotic therapy should be administered by the end of hospital day two to reduce stroke mortality and morbidity as long as no contraindications exist.

Healthcare organizations that track early antithrombotic therapy for internal quality improvement purposes have seen an improvement in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
In October, 2009, The Joint Commission added the stroke (STK) measure set as a new core measure option to meet performance measure requirements for Joint Commission hospital accreditation purposes. Below is the specified level of analysis for STK-5 beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 1955 denominator cases; 1875 numerator cases; 49 hospitals; 0.95908 national aggregate rate; 0.95947 mean of hospital rates; 0.10574 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.94949 50th percentile rate/median rate; 0.85165 25th percentile rate/lower quartile; and, 0.66667 10th percentile rate.

CY 2010: 18,728 denominator cases; 18,223 numerator cases; 137 hospitals; 0.97304 national aggregate rate; 0.95924 mean of hospital rates; 0.08189 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.98361 50th percentile rate/median rate; 0.96183 25th percentile rate/lower quartile; and, 0.90909 10th percentile rate.

CY 2011: 22,328 denominator cases; 21,814 numerator cases; 157 hospitals; 0.97698 national aggregate rate; 0.95933 mean of hospital rates; 0.10243 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.98667 50th percentile rate/median rate; 0.96053 25th percentile rate/lower quartile; and, 0.928 10th percentile rate.

CY 2012: 22,258 denominator cases; 21,866 numerator cases; 158 hospitals; 0.98239 national aggregate rate; 0.98042 mean of hospital rates; 0.03174 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99396 50th percentile rate/median rate; 0.97143 25th percentile rate/lower quartile; and, 0.95098 10th percentile rate.

CY 2013: 33,651 denominator cases; 33,133 numerator cases; 263 hospitals; 0.98461 national aggregate rate; 0.97976 mean of hospital rates; 0.04874 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99422 50th percentile rate/median rate; 0.9781 25th percentile rate/lower quartile; and, 0.95745 10th percentile rate.

CY 2014: 162,275 denominator cases; 159.848 numerator cases; 1300 hospitals; 0.98504 national aggregate rate; 0.97832 mean of hospital rates; 0.07287 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99261 50th percentile rate/median rate; 0.97832 25th percentile rate/lower quartile; and, 0.95707 10th percentile rate.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
Not applicable

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al. 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

In the national REGARDS cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years of age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age : RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer, et al., 2010).

In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).
In terms of early antithrombotic administration, no significant disparities were noted across studies. Recent cross-sectional studies from nationwide health surveys suggest that once patients have a stroke, the use of most risk-reduction strategies does not vary significantly by race. An analysis of 1045 veteran patients found the use of aspirin similar for blacks or African Americans and whites, although the intervention was underutilized with aspirin administration in approximately 74% for both groups (Ross and Bravata, 2009). Similarly, data from the Centers for Disease Control (CDC) Behavioral Risk Factor Surveillance System (BRFSS) showed that black or African American stroke survivors were no less likely to be prescribed aspirin than whites.

Since the last endorsement date, several studies have reported results similar to those from BRFSS in 2009. Schwamm and colleagues (2010) found significant and important differences in quality of care related to early antithrombotics when multivariate models were constructed adjusting for patient-level characteristics only. Race/ethnicity Black versus White: unadjusted OR 0.96 [95% CI 0.92-1.00]; adjusted for patient characteristics OR 0.90 [95% CI 0.86-0.94]; adjusted for patient and hospital characteristics OR 0.97 [95% CI 0.91-1.02]. Race/ethnicity Hispanic versus White: unadjusted OR 0.95 [95% CI 0.88-1.02]; adjusted for patient characteristics OR 0.89 [95% CI 0.83-0.96]; adjusted for patient and hospital characteristics OR 0.96 [95% CI 0.88-1.05]. Total N=271,769; All n 94.23%; White 94.28%; Black 94.08%; Hispanic 93.98%.

A more recently published study (Qian F, et al, 2013) from GWTG found that non-Hispanic black patients were less likely to receive early antithrombotics when compared to other race/ethnicity groups. Using patient data (n=200,900) from the American Heart Association/American Stroke Association Get With The Guidelines (GWTG)-Stroke program from April 2003 through December 2008, the following performance measure rates were reported for early antithrombotic therapy: non-Hispanic White (n=170,694) 94.9%; non-Hispanic Black (n=20,514) 94.0%; Hispanic (n=6632) 94.3%; and non-Hispanic Asian American (n=3060) 94.8% Data from the Paul Coverdell National Acute Stroke Registry (PCNASR) for early antithrombotic therapy were similar (n=58,823): White 96.5%; Other Race 95.6% (Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, 2014).

• Ambriz EH, Woodard LD, Kressin NR, Petersen LA. Use of smoking cessation interventions and aspirin for secondary prevention: are there racial disparities? Am J Med Qual. 2004; 19:166-171.
• Bhattacharya P, Mada F, Salowich-Palm L, Hinton S, Millis S, Watson SR, Chaturvedi S, Rajamani K. Department of neurology and Stroke Program, Wayne State University School of Medicine, Detroit, Michigan. Are racial disparities in stroke care still prevalent in certified stroke centers? J Stroke Cerebrovasc Dis. 2011.
• Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, Yan T, Escarce JJ. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-8.
• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, Griffith P, Gorelick PB, Howard G, Leira EC, Morgenstern LB, Ovbiagele B, Peterson E, Rosamond W, Trimble B, Valderrama AL, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• Gillum RF,Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 200;42(11):3294-6.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-75.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Jacobs BS, Birbeck G, Mullard AJ, Hickenbottom S, Kothari R, Roberts S, Reeves MJ. Quality of hospital care in African Americans and white patients with ischemic stroke and TIA. Neurology.2006;66:809-14.
• Karve S, Balkrishnan R, Seiber E Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.

• Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, Kerber KA, Skolarus LE, Meurer WJ, Burke JF, Adelman EE, Baek J, Lisabeth LD. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol. 2013;74:778–785.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Ross JS, Bravata DM. Use of stroke secondary prevention services: are there disparities in care? Stroke. 2009;40:1811-1819.
• Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Schwamm LH, Syed FA, Reeves MJ, Smith EE, Saver JL, Messe S, Bhatt DL, Grau-Sepulveda MV, Peterson ED, Fonarow GC. Temporal trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With the Guidelines-Stroke hospitals. Circ Cardiovasc Quality Outcomes. 2013;6:543-549.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States remains attributable to stroke. More women than men die of stroke each year. Women accounted for almost 60% of US stroke deaths (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, et al., 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 (P<0.05).168 (US Burden of Disease Collaborators, 2013) . Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

Antithrombotic agents significantly reduce the incidence of a recurrent vascular event after a stroke. Among these agents, aspirin has been the drug most widely studied. The International Stroke Trial demonstrated that antithrombotic administration, specifically aspirin, within the first 48 hours after stroke, significantly reduced the risk recurrent ischemic stroke and death (11.3% vs. 12.4%) in

the first 14 days following the event. In conjunction with this, findings from the Chinese Acute Stroke Trial indicate that aspirin produces a modest reduction of approximately 10 deaths per 1000 during the first few weeks. Both trials recommend that aspirin should be given as soon as possible after the onset of stroke symptoms. It appears that the primary benefits of aspirin are due to early reduction in recurrent stroke rather than limitation of neurological deficits of the first stroke. For early antithrombotic therapy, oral administration of aspirin (initial dose is 325 mg) within 24 to 48 hours after stroke onset remains the current clinical guideline recommendation for treatment of most patients (Jauch, E. C., et al., 2013; Sandercock, P. A., et. al., 2014).

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

• Adams HP, del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, Grubb RL, Higashida RT, Jauch EC, Kidwell C, Lyden PD, Morgenstern LB, Qureshi AI, Rosenwasser RH, Scott PA, Wijdicks E. Guidelines for the Early Management of Adults with Ischemic Stroke: A Guideline From the American Heart Association/American Stroke Association Stroke Council, Clinical Cardiology Council, Cardiovascular Radiology and Intervention Council, and the Atherosclerotic Peripheral Vascular Disease and Quality of Care Outcomes in Research Interdisciplinary Working Groups. Stroke. 2007;38:1655-1711.
• Albers GW, Amarenco P, Easton JD, Sacco RL, Teal P. Antithrombotic and Thrombolytic Therapy for Ischemic Stroke. Chest Vol. 119, 2001: 300-320.
• Brott TG, Clark WM, Grotta JC, et al. Stroke the first hours. Guidelines for acute treatment. Consensus Statement. National Stroke Association. 2000.
• Chen ZM, Sandercock P, on behalf of the AntiThrombotic Trialists Collaboration (ATT). Indications for early aspirin use in acute stroke: a systematic overview of over 40,000 randomised patients. Cerebrovasc Dis. 1998;8(suppl 4):38.
• Chen ZM, Sandercock P, Pan HC, et al. Indications for early aspirin use in acute ischemic stroke: a combined analysis of 40,000 randomized patients from the Chinese acute stroke trial and the international stroke trial. On behalf of the CAST and IST collaborative groups, Stroke 2000;31:1240-1249.
• Coull BM, Williams LS, Goldstein LB, et al. Anticoagulants and Antiplatelet Agents in Acute Ischemic Stroke. Report of the Joint Stroke Guideline Development Committee of the American Academy of Neurology and the American Stroke Association (a Division of the American Heart Association) Stroke. 2002;33:1934 -1942.
• George M, Xin T, McGruder H, Yoon P, Rosamond W, Winquis, A, Hinchey J, Wall H, Pandey D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Guyatt GH, Akl EA, Crowther M, Gutterman DD, Schu?nemann HJ, and for the American College of Chest Physicians Antithrombotic Therapy and Preveniton of Thrombosis Panel. Chest. 2012;141: 33S.
• Guideline on the Use of Aspirin as Secondary Prophylaxis for Vascular Disease in Primary Care, Centre for Health Services Research University of Newcastle upon Tyne, & Centre for Health Economics of York, 1998.
• Jauch EC, Saver JL, Adams HP, Bruno A, Connors JJ, Demaerschalk BM, Khatri P, McMullan PW, Qureshi AI, Rosenfield K, Scott PA, summers DR, Wang DZ, Wintermark M, Yonas, H. Guidelines for the early management of patients with acute ischemic stroke: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44:893-899.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Sandercock P, Collins R, Counsell C, Farrell B, Peto R, Slattery J, Warlow C, International Stroke Trial Collaborative Group: The International Stroke Trial (IST): A randomised trial of aspirin, subcutaneous heparin, both, or neither among 19,435 patients with acute ischaemic stroke. Lancet. 1997;349:1569-1581.
• Sandercock P, Counsell C, Tseng MC, Cecconi E. Oral antiplatelet therapy for acute ischaemic stroke. The Cochrane Database of Systematic Reviews. 2014;3:CD000029.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Appendix_A.1-635878644173852080.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
There have been no significant changes to the STK-05: Antithrombotic Therapy By End of Hospital Day 2 measure since the last endorsement date.

All ICD-9-CM diagnosis codes and ICD-9-CM procedure codes were converted to ICD-10-CM diagnosis and ICD-10-PCS procedure codes throughout the measure specifications. The Department of Health and Human Services (HHS) mandated that all entities covered by the Health Insurance Portability and Accountability Act (HIPAA) must transition to a new set of codes for electronic health care transactions on October 1, 2015.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Ischemic stroke patients who had antithrombotic therapy administered by
end of hospital day two.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
One data element is used to calculate the numerator:
• Antithrombotic Therapy Administered by End of hospital Day 2 – Documentation that antithrombotic therapy is administered by

the end of hospital day 2. Allowable values: Yes, No/UTD or unable to determine from medical record documentation.

Patients are eligible for the numerator population when the allowable value equals "yes" for the data element.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Ischemic stroke patients

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Ten data elements are used to calculate the denominator:

1. Admission Date – The month, day and year of admission to acute inpatient care.
2. Arrival Date – The earliest documented month, day, and year the patient arrived at the hospital.
3. Birthdate - The month, day and year the patient was born.
4. Clinical Trial - Documentation that during this hospital stay the patient was enrolled in a clinical trial in which patients with stroke were being studied. Allowable values: Yes or No/UTD.
5. Comfort Measures Only – The earliest day the physician/APN/PA documented comfort measures only after hospital arrival. Allowable values: 1 (Day 0 or 1); 2 (Day 2 or after); 3 (Timing Unclear); 4 (Not Documented/UTD).
6. Discharge Date – The month day and year the patient was discharged from acute care, left against medical advice or expired during the stay.
7. Elective Carotid Intervention – Documentation demonstrates that the current admission is solely for the performance of an elective carotid intervention (e.g., elective carotid endarterectomy, angioplasty, carotid stenting).
Allowable values: Yes or No/UTD.
8. ICD-10-CM Principal Diagnosis Code - The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.
9. IV OR IA Thrombolytic (t-PA) Therapy Administered at this Hospital or within 24 Hours Prior to Arrival – Documentation demonstrates that the patient received intravenous (IV) or intra-arterial (IA) thrombolytic therapy (t-PA) at this hospital or within 24 hours prior to arrival.
Allowable values: Yes or No/UTD.
10. Reason for Not Administering Antithrombotic Therapy By End of Hospital Day 2 – Physician/APN/PA or pharmacist documentation of a reason for not administering antithrombotic therapy by end of hospital day 2.
Allowable values: Yes or No/UTD.

Population: Discharges with ICD-10-CM Principal Diagnosis Code for ischemic stroke as defined in Appendix A, Table 8.1

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
- Less than 18 years of age
- Duration of Stay < 2 days
- Length of Stay > 120 days
- Comfort measures only documented on the day of or day after hospital arrival
- Enrolled in clinical trials related to stroke
- Admitted for elective carotid intervention
- IV OR IA thrombolytic therapy administered at this hospital or within 24 hours prior to arrival
- Documented reason for not administering antithrombotic therapy by end of hospital day 2

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
- The patient age in years is equal to the Discharge Date minus the Birthdate. Patients less than 18 years are excluded.

- The Duration of Stay (in days) is equal to the Discharge Date minus the Arrival Date. If the Duration of Stay is less than 2 days, the

patient is excluded.

• The Length of Stay (LOS) in days is equal to the Discharge Date minus the Admission Date. If the LOS is greater than 120 days, the patient is excluded.

• Patients with Comfort Measures Only allowable value of 1 (Day 0 or 1) are excluded.

• Patients are excluded if "Yes" is selected for Clinical Trial.

• Patients are excluded with ICD-10-PCS procedure codes for carotid intervention procedures as identified in Appendix A, Table 8.3, if medical record documentation states that the patient was admitted for the elective performance of this procedure.

• Patients are excluded if "Yes" is selected for IV (intravenous) or IA (intra- arterial)Thrombolytic Therapy (t-PA)Administered at This Hospital or Within 24 Hours Prior to Arrival.

• Patients are excluded if "Yes" is selected for Reason For Not Administering Antithrombotic Therapy By End of Hospital Day 2.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable, the measure is not stratified.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not applicable

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
1. Start processing. Run cases that are included in the Stroke (STK) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check ICD-10-CM Principal Diagnosis Code
a. If the ICD-10-CM Principal Diagnosis Code is not on Table 8.1, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.

b. If the ICD-10-CM Principal Diagnosis Code is on Table 8.1, continue processing and proceed to Comfort Measures Only.

3. Check Comfort Measures Only
a. If Comfort Measures Only is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Comfort Measures Only equals 1, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Comfort Measures Only equals 2, 3, or 4, continue processing and proceed to Clinical Trial.

4. Check Clinical Trial
a. If Clinical Trial is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Clinical Trial equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Clinical Trial equals No, continue processing and proceed to Elective Carotid Intervention.

5. Check admitted for Elective Carotid Intervention
a. If Elective Carotid Intervention is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Elective Carotid Intervention equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Elective Carotid Intervention equals No, continue processing and proceed to Arrival Date.

6. Check Arrival Date
a. If the Arrival Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If the Arrival Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
c. If the Arrival Date equals a Non-Unable To Determine (non-UTD) Value, continue processing and proceed to Duration of Stay calculation.

7. Calculate the Duration of Stay. The Duration of Stay, in days, is equal to the Discharge Date minus the Arrival Date.

8. Check Duration of Stay
a. If the Duration of Stay is greater than or equal to zero and less than 2, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
b. If the Duration of Stay is greater than or equal to 2, continue processing and proceed to IV or IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival.

9. Check IV or IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival
a. If IV or IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If IV or IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If IV or IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival equals No, continue processing and proceed to Antithrombotic Therapy Administered By End of Hospital Day 2.

10. Check Antithrombotic Therapy Administered By End of Hospital Day 2
a. If Antithrombotic Therapy Administered By End of Hospital Day 2 is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Antithrombotic Therapy Administered By End of Hospital Day 2 equals Yes, the case will proceed to a Measure Category Assignment of E and will be in the Numerator Population. Stop processing.
c. If Antithrombotic Therapy Administered By End of Hospital Day 2 equals No, continue processing and check Reason for Not Administering Antithrombotic Therapy By End of Hospital Day 2.

11. Check Reason for Not Administering Antithrombotic Therapy By End of Hospital Day 2
a. If Reason for Not Administering Antithrombotic Therapy By End of Hospital Day 2 is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Reason for Not Administering Antithrombotic Therapy By End of Hospital Day 2 equals Yes, the case will proceed to a Measure

Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Reason for Not Administering Antithrombotic Therapy By End of Hospital Day 2 equals No, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter for the measure set cannot sample.

Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

Quarterly Sampling
The Quarterly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing quarterly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If "N" >/= 900, then "n" 180
If "N" 226-899, then "n" 20% of Initial Patient Population size
If "N" 45-225, then "n" 45
If "N" 6-44, No sampling; 100% Initial Patient Population required
If "N" 0-5, Submission of patient level data is not required; if submission occurs, 100% Initial Patient Population required

Monthly Sampling
The Monthly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing monthly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If "N >/= 300, then "n" 60
If "N 76-299, then "n" 20% of Initial Patient Population size
If "N 15-75, then "n" 15
If "N < 15, No sampling; 100% Initial Patient Population required

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Each data element in the data dictionary includes suggested data sources. The data are collected using contracted Performance

Measurement Systems (vendors) that develop data collection tools based on the measure specifications. The tools are verified and tested by Joint Commission staff to confirm the accuracy and conformance of the data collection tool with the measure specifications. The vendor may not offer the measure set to hospitals until verification has been passed.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0438_MeasureTesting_MSF5.0_Data-635905393038485759.doc

NQF #: 0438 NQF Project: Neurology Project

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009.. Demographics of organizations collecting and reporting data on this measure are as follows:
170 health care organizations representing various hospital demographics:
17 For Profit; 144 Not for Profit; 9 Government
62 >=300 beds; 71 100-299 beds; 37 <100 beds
142 Urban; 28 Rural
22 Teaching; 148 Non-teaching
States represented in this data collection effort include: AL, AR, AZ, CA, CT, FL, GA, IA, ID, IL, IN, KY, LA, MA, MD, MI, MN, MO, MS, NC, NE, NJ, NM, NY, OH, PA, PR, SC, SD, TN, TX, VA, WA, WI
15 performance measurement systems are used for data transmission to The Joint Commission.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with monthly data transmission to The Joint Commission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of approximately 900 patient records identified at 30 pilot sites over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rates of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify potential measure enhancements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures.) The following is a list of the major tests conducted on the submitted data for this measure:

- Transmission of complete data
- Usage of data received (to understand if the healthcare organization provides the relevant service to treat the relevant population)
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity

- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors is as follows:
- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Data element agreement rates reported to The Joint Commission for the time period of one year (4Q2010 – 3Q2011) have shown an overall agreement rate of 97.2%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-5.

Data Elements	total'n'numerator	total'n'denominator	rate
Antithrombotic Therapy Administered By End of Hospital Day 2	426	449	95.0%
Clinical Trial	712	714	99.7%
Comfort Measures Only	709	714	99.3%
Elective Carotid Intervention	711	714	99.6%
IV OR IA Thrombolytic Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival	464	476	97.5%
Reason for Not Administering Antithrombotic Therapy by End of Hospital Day 2	44	47	93.6%

These agreement rates are considered to be well within acceptable levels.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* are consistent with the evidence cited in support of the measure focus (*criterion 1c)* and identify any differences from the evidence:
This measure focuses on determining the proportion of ischemic stroke patients who had antithrombotic therapy administered by end of hospital day two. The literature supports that antithrombotic administration by end of hospital day two significantly reduces the risk of ischemic stroke recurrence in the initial weeks following the first stroke with minimal risk of a major bleed. The measure specifications are consistent with AHA guideline recommendations for aspirin administration within the first 24 to 48 hours following ischemic stroke onset. This measure excludes patients less than 18 years of age. Also excluded from the measure are patients who have a length of stay (LOS) of less than two days or more than 120 days, those who are enrolled in a clinical trial for stroke or for whom comfort measures only have been ordered on the day of or the day after arrival. These exclusions are not addressed in the literature, but are included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

The measure specifications differ from guideline recommendations by excluding patients admitted for Elective Carotid Intervention. Operationally, patients with a planned admission to the hospital for an elective carotid intervention which restores blood flow to the brain and prevents stroke, such as carotid endarterectomy or carotid stenting, are excluded from the measure. Additionally, patients who receive IV or IA Thrombolytic Therapy at this hospital or within 24 hours prior to arrival are excluded since antithrombotic administration following thrombolytic therapy within this timeframe may increase the risk of hemorrhage, and is therefore, inadvisable.

The computation used to determine the two-day timeframe for this measure is the key difference between the measure specifications and guideline recommendations. The measure specifications use days (i.e., day 1 = day of hospital arrival and day 2 = day after arrival) rather than hours and minutes (i.e., 24 to 48 hours) to compute the end of hospital day 2. This computation benefits both the patient and healthcare organization. One benefit is that it eases the burden of data abstraction. Depending on the patient's time of arrival, there may be less time for administration of antithrombotic therapy for some cases; however, early antithrombotic administration post-stroke is the goal and consistent with the guideline time range recommendation of 24 to 48 hours.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:
At the time this measure was originally tested measure validity was assessed via survey and focus groups of hospitals participating in the pilot test. All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant.

Since the measure has been in national use, continued validity of the measure has been determined through analysis of feedback from measure users. The Joint Commission provides a web-based application with which measure users can provide feedback regarding appropriateness of measure specifications, request clarification of specifications, and/or provide other comments pertinent to the measure. This feedback is systematically, continually, reviewed in order to identify trends and to identify areas of the measure specifications that require clarification or revision. Additionally, Joint Commission staff continually monitors the national literature and environment in order to assess continued validity of this measure. An expert technical advisory panel (TAP) meets on a quarterly basis in order to assess continued validity of this measure. And finally, the conversion from ICD-9-CM diagnosis codes to ICD-10-CM diagnosis codes has been completed and reviewed by the Technical Advisory Panel for validity. The panel has determined that the intent of the measure has not changed as a result of the conversion. The crosswalk will also be posted in a future version of the specifications manual for public comment, and results of feedback will be reviewed and incorporated into the measure specifications where indicated.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:
Analysis of feedback obtained via our automated feedback system reveals 72 submissions regarding specifications for this measure over the past year. Predominant themes of these submissions involved questions regarding clarification of the data element Reason for Not Prescribing Antithrombotic Therapy By End of Hospital Day 2, with respect to both acceptable reasons and the timeframe for documentation. Additional notes for abstractors were added to the data element for clarification. Inquiries about new anticoagulants dabigatran and rivaroxaban pertaining to inclusion of patients taking these medications in the numerator were also received. The medication Table 8.2 for the data element Antithrombotic Therapy By End of Hospital Day 2 has been updated to include these medications.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:
Measure exclusions that were not derived directly from the evidence are presented below. Please note that these are population exclusions that are necessary to ensure consistency in all measures in this 8 measure set.
These exclusions were analyzed for frequency of occurrence. An issue that is of great concern to users of this measure is that due to the presence of exceptions to the measure, attainment of a 100% measure rate is not possible. Because of the role of this measure in the current Joint Commission accreditation process this is especially troubling to measure users. This concern is the basis for a number of the non-evidence-based exclusions to these measures. The following measure exclusions that were not derived directly from the evidence are as follows:

1. Patients < 18 years old
2. Patients who have a Duration of Stay less than 2 days
3. Patients who have a length of stay (LOS) greater than 120 days
4. Patients with Comfort Measures Only documented on day of or after arrival
5. Patients enrolled in clinical trials
6. Patients admitted for Elective Carotid Intervention

7. Patients with a documented Reason For Not Administering Antithrombotic Therapy By End of Hospital Day 2

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
N=39,812
1. Patients < 18 years old = 0%
2. Patients who have a Duration of Stay less than 2 days = 8.43%
3. Patients who have a length of stay (LOS) greater than 120 days = 0%
4. Patients with Comfort Measures Only documented = 4.99%
5. Patients enrolled in clinical trials = 0.39%
6. Patients admitted for Elective Carotid Intervention = 9.96%
7. Patients with a documented Reason For Not Administering Antithrombotic Therapy By End of Hospital Day 2 = 2.27%

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Not Applicable

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:
Not Applicable

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:
Not Applicable

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:
Not Applicable

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:
The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of an HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
STK-5 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10th Percentile
3Q2011	154	0.97045	0.0735	1	1	1	0.9697	0.92857
2Q2011	155	0.96839	0.08678	1	1	1	0.97143	0.93333
1Q2011	159	0.95328	0.12687	1	1	1	0.95833	0.91304
4Q2010	135	0.96697	0.104	1	1	1	0.96667	0.93023
3Q2010	129	0.95443	0.12216	1	1	1	0.96226	0.88889
2Q2010	122	0.95333	0.12219	1	1	1	0.95556	0.90625
1Q2010	100	0.9521	0.12692	1	1	1	0.96	0.87298
4Q2009	49	0.95947	0.10574	1	1	0.98519	0.96875	0.93478

It should be noted that since data collection on this measure is completely voluntary for both The Joint Commission and PCNASR, the hospitals reporting on this measure are self-selected, and therefore, presumably have a particular interest and concern for improving stroke care. For this reason, the measure results demonstrated by this group most likely significantly overstates the rate for the population of all health care organizations.

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Multiple data sources are not used for this measure.

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:
Not applicable

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:
Not applicable

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: The measure is not stratified for disparities.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:
There are no plans to stratify the measure. The evidence does not support that minorities are less likely to receive antithrombotic therapy by end of hospital day two. The Joint Commission does not currently capture gender-specific data, or data elements for race or ethnicity because these data elements have not been shown to be reliably collectable due to the fact that no national standardized definitions exist for these data elements. Also, not all hospitals collect race and ethnicity. In the future, it may be feasible for The Joint Commission to explore how race and ethnicity and other relevant disparity data, might be collected reliably in the future.

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?
(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

NEW (since 2012 endorsement): Data for Empirical Testing

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The data used to measure validity consists of one year data (third, and fourth quarter 2014 and first and second quarter for 2015) extracted from The Joint Commission warehouse. The hospitals selection was based to those hospitals that reported 12 months of data and had 30 or more denominator cases for the year.

Measure convergent and criterion validity was assessed using hospitals patient level data from The Joint commission warehouse. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous face validity testing. The patient population was comprised of all patients discharged with an ICD-9-CM Principal Diagnosis Code for stroke as defined in Appendix A, Table 8.1 or Table 8.2.

Conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications. The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015.

We conducted a secondary analysis to help interpret results; correlation with other stroke process measures.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Descriptive statistics for the STK measures:
N=1,318 hospitals
n= 2, 206,379 patients records

STK-5
Median: 100%
Percentile 10%: 98%
Percentile 25%: 99%
Percentile 75%: 100%
Percentile 90%: 100%

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations							
	STK_1	STK_2	STK_3	STK_4	STK_5	STK_6	STK_10
STK_1	1.00000 1318	0.55423 <.0001 1318	0.30311 <.0001 1302	0.37009 <.0001 1144	0.52785 <.0001 1318	0.53082 <.0001 1318	0.45291 <.0001 1318
STK_2	0.55423 <.0001 1318	1.00000 1318	0.37566 <.0001 1302	0.28169 <.0001 1144	0.53355 <.0001 1318	0.47326 <.0001 1318	0.29578 <.0001 1318
STK_3	0.30311 <.0001 1302	0.37566 <.0001 1302	1.00000 1302	0.25551 <.0001 1137	0.25665 <.0001 1302	0.28198 <.0001 1302	0.15819 <.0001 1302
STK_4	0.37009 <.0001 1144	0.28169 <.0001 1144	0.25551 <.0001 1137	1.00000 1144	0.22516 <.0001 1144	0.43717 <.0001 1144	0.41076 <.0001 1144
STK_5	0.52785 <.0001 1318	0.53355 <.0001 1318	0.25665 <.0001 1302	0.22516 <.0001 1144	1.00000 1318	0.35302 <.0001 1318	0.39079 <.0001 1318
STK_6	0.53082 <.0001 1318	0.47326 <.0001 1318	0.28198 <.0001 1302	0.43717 <.0001 1144	0.35302 <.0001 1318	1.00000 1318	0.45420 <.0001 1318
STK_10	0.45291 <.0001 1318	0.29578 <.0001 1318	0.15819 <.0001 1302	0.41076 <.0001 1144	0.39079 <.0001 1318	0.45420 <.0001 1318	1.00000 1318

The correlations of the stroke measures are all positively correlated and statistically significant indicating that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)
Overall, the positive inter-correlations indicates convergent validity of the measures.
STK-05 is positively correlated with other stroke evidence-based processes of care.

2b3. EXCLUSIONS ANALYSIS .
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patients who have a Duration of Stay less than 2 days
- Patient who have a Length of Stay greater than 120 days

- Patients with *Comfort Measures* documented on day of or day after arrival
- Patient enrolled in a Clinical Trial
- Patients admitted for *Elective Carotid Intervention*
- Patients discharged prior to the end of hospital day 2
- Patients with *IV **OR** IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival*
- Patients with a documented *Reason for Not Administering Antithrombotic Therapy by End of Hospital Day 2*

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)
There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are given below.

Exclusion: *Comfort Measures Only* documented on day of or day after arrival
Overall Occurrence n = 28,126
Overall Occurrence Percentage 4.46%
Minimum 0.31%
10th Percentile: 1.67%
Median: 4.40%
90th Percentile: 9.09%
Maximum: 30.8%

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%
10th Percentile: 0.24%
Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for *Elective Carotid Intervention*
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.5%
90th Percentile: 23.5%
Maximum: 95.2%

Exclusion: Patients with *IV **OR** IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival*
Overall Occurrence n = 25,354
Overall Occurrence Percentage 8.04%
Minimum: 0.25%
10th Percentile: 2.1%
Median: 6.72%
90th Percentile: 13.5%
Maximum: 47.4%

Exclusion: Patients with a documented *Reason for Not Administering Antithrombotic Therapy by End of Hospital Day 2*
Overall Occurrence n = 22,404

Overall Occurrence Percentage 7.11%
Minimum: 0.351%
10th Percentile: 1.48%
Median: 4.16%
90th Percentile: 14.9%
Maximum: 76.9%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The median frequency of exclusions range from low to moderate. The distribution of exclusions across hospitals is generally narrow indicating that the occurrence is random and likely would not bias performance results, although the percentage of patients excluded may differ depending on whether the hospital was a stroke center.

For criterion validity, it is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a Length of Stay greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Patients who have a Duration of Stay less than 2
Rationale: Inclusion of patients with a duration of stay shorter than 2 days may falsely increase the denominator population.

Patients with *Comfort Measures Only* documented on day of or day after arrival
Rationale: It is inappropriate to treat such patients beyond those measures necessary for comfort.

Patients enrolled in a Clinical Trial
Rationale: Antithrombotic therapy as specified by this measure may compromise a clinical trial.

Patients admitted for *Elective Carotid Intervention*
Rationale: Patients who are not admitted for treatment of an acute stroke may not require antithrombotic therapy

Patients with *IV **OR** IA Thrombolytic (t-PA) Therapy Administered at This Hospital or Within 24 Hours Prior to Arrival*
Rationale: It is inappropriate to administer antithrombotic therapy to patients who have received IV or IA thrombolytic therapy within 24 hours

Patients with a documented Reason For Not Administering Antithrombotic Therapy By End of Hospital Day 2
Rationale: It is inappropriate to treat patients who have a documented reason or contraindication to antithrombotic therapy

Patients discharged prior to the end of hospital day 2
Rationale: Inclusion of patients discharged prior to the end of hospital day 2 may falsely increase the denominator population.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other

If other: Data element allowable values are selected, either manually or electronically, from clinical and coded data available in medical record documentation. All medical record documentation is used in the abstraction process. Vendor data collection tools are used to import data elements needed for measure rate calculation.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract the electronic version of this measure, so continues to offer this chart-abstracted version which allows for data capture from unstructured data fields. All data elements needed to compute the STK-5 performance measure score have been retooled for capture from electronic sources. Annual updates are performed to match the eCQM specifications to the current version of the chart-abstracted specifications.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

At the present time, hospitals using this performance measure generally collect measure data via manual review of the paper medical record, the EHR or a combination of both. Collected data are submitted to The Joint Commission on a quarterly basis, by way of contracted performance measurement system vendors, as described previously. Specifications for this measure are freely available to anyone who wishes to use the measure. Feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As described above, as feedback from measure users has indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

There are no fees or licensing requirements to use The Joint Commission performance measures, all of which are in the public

domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Quality Check® http://www.qualitycheck.org/consumer/searchQCR.aspx Hospital Compare https://www.medicare.gov/hospitalcompare/search.html Public Health/Disease Surveillance Paul Coverdell National Acute Stroke Registry http://www.cdc.gov/dhdsp/programs/stroke_registry.htm Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program http://www.jointcommission.org/ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Annual Report – Improving America's Hospitals http://www.jointcommission.org/annualreport.aspx Quality Improvement (Internal to the specific organization) Disease-Specific Care Certification for Comprehensive Stroke http://www.jointcommission.org/certification/dsc_home.aspx http://www.jointcommission.org/certification/dsc_home.aspx Disease-Specific Care Certification for Primary Stroke Centers

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

• Name of program and sponsor: Quality Check®; The Joint Commission

• Purpose: A public website that allows consumers to: search for accredited and certified organizations by city and state, by name or

by zip code (up to 250 miles); find organizations by type of service provided within a geographic area; download free hospital performance measure results; and, print a list of Joint Commission certified disease-specific care programs and health care staffing firms.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

• Name of program and sponsor: Hospital Compare; Centers for Medicare & Medicaid Services
• Purpose: A public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4000+ medicare-certified hospitals (2015)

• Name of program and sponsor: Paul Coverdell National Acute Stroke Registry; Centers for Disease Control and Prevention
• Purpose: A national registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.
• Geographic area and number and percentage of accountable entities and patients included: 11 states; 403 hospitals (CDC, 2014)

• Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
• Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 hospitals (2015)

• Name of program and sponsor: Annual Report-Improving America's Hospitals; The Joint Commission
• Purpose: The annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

• Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes that specific capabilities of hospitals that treat the most complex stroke cases.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

• Name of program and sponsor Disease-Specific Care Certification for Primary Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
• Geographic area and number and percentage of accountable entities and patients included Nationwide; 1079 hospitals

• Name of program and sponsor Hospital Accreditation Program; The Joint Commission
• Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
• Geographic area and number and percentage of accountable entities and patients included Nationwide; 3300 Joint Commission-accredited hospitals (2014)

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data*

aggregation and reporting.)
Not applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

•Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)

The rate of early antithrombotic therapy has steadily increased over the past five years based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 98.5%. A modest gap of ~4% still exists for the 10th percentile of hospitals. This trend is consistent with increases in the rate of antithrombotic therapy at discharge published by GWTG-Stroke and PCNASR. Significant and important differences between races in rates of early antithrombotic therapy persist for Hispanics and non-Hispanic blacks (Schwamm, 2010, Qian, 3013; CDC 2014).

•Geographic area and number and percentage of accountable entities and patients included Nationwide; 1300 hospitals; 162,275 patients (The Joint Commission, 2014)

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
Omission of early antithrombotic therapy when indicated may result in poorer neurological outcomes for stroke patients. The timeframe specified for antithrombotic administration is the day of or day after hospital arrival. The two-day timeframe is aligned with the guideline recommendation that aspirin should be administered within 24 to 48 hours of stroke onset. A two-day computation is used rather than abstracting time in precise minutes and hours to ease abstraction burden. Antithrombotic therapy must be administered by 2359 on the day after hospital arrival (day two), or a reason for not administering documented; however, this methodology for abstraction may result in a fall-out from the measure if the patient arrives late in the day on day one and antithrombotic therapy is administered within 48 hours of hospital arrival.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0068 : Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet

0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy
0435 : STK 02: Discharged on Antithrombotic Therapy

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0325: Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy – American Academy of Neurology

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Measures 0435 Discharged on Antithrombotic Therapy is the second (STK-2) measure in The Joint Commission stroke core measure set and also targets the ischemic stroke population; however, the timeframe for antithrombotic administration is different in this measure than STK-5. STK-2 focuses on hospital discharge and the prescription of antithrombotic medications at that time. All common data elements are completely aligned between the two measures. Measure 0068 is a physician performance measure and thus a different level of measurement. Measure 0068 encompasses a different target population, specifically patients with ischemic vascular disease who were discharged alive for acute myocardial infarction (AMI), coronary artery bypass graft (CABG) or percutaneous coronary interventions (PCI). Both of these measures evaluate physician practice as opposed to hospital processes

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not Applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Ann, Watt, awatt@jointcommission.org, 630-792-5944-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Karen, Kolbusz, kkolbusz@jointcommission.org, 630-792-5931-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel (TAP) is to provide advisory oversight in literature review, measure content and maintenance of the specifications.
Members are:

Harold P. Adams, Jr., MD
University of Iowa Health Care
Iowa City, IA

Mark J. Alberts, MD
University of Texas Southwestern
Dallas, TX

Anne W. Alexandrov, RN
Health Outcomes Institute
Fountain Hills, AZ

Nadine Allyn, RD
American Heart Association
Dallas, Texas

Mary G. George, MD
Centers for Disease Control and Prevention
Atlanta, GA

Martin Gizzi, MD
Legacy Health System
Portland, OR

Marianna Gorbaty
Mathematica Policy Research, Inc.
Washington, DC

Judith Hinchey, MD
Tuft's Medical School
Steward St. Elizabeth's Medical Center
Boston, MA

Edward C. Jauch, MD
Medical University of South Carolina
Charleston, SC 2946

Irene Katzan, MD
Cleveland Clinic
Cleveland, OH

Kimberly E. Levasseur-Franklin, PharmD
Northwestern Memorial Hospital
Chicago, IL

Kathy Morrison, RN
Penn State Hershey Medical Center
Hershey, PA

Marilyn M. Rymer, MD
University of Kansas
Kansas City, MO

Jeffrey Saver, MD
UCLA Medical Center
Los Angeles, CA

Lee H. Schwamm, MD
Harvard Medical School
Massachusetts General Hospital
Boston, MA

Deborah, Summers, RN
St. Luke's Brain and Stroke Institute
Kansas City, MO

Osama O. Zaidat, MD
St.Vincent Mercy Hospital
Toledo, OH

Richard D. Zorowitz, MD
Medstar National Rehabilitation Network
Washington, DC

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2009

Ad.3 **Month and Year of most recent revision:** 10, 2015

Ad.4 **What is your frequency for review/update of this measure?** Biannual

Ad.5 **When is the next scheduled review/update for this measure?** 07, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0439
Measure Title: STK-06: Discharged on Statin Medication
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of ischemic stroke patients who are prescribed a statin medication at hospital discharge.

This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-4: Thrombolytic Therapy, STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-8: Stroke Education, and STK-10: Assessed for Rehabilitation) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs.
Developer Rationale: Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Recent guidelines recommend high-intensity statin therapy for patients with ischemic stroke unless contraindicated or patient characteristics require dosage modification.

Healthcare organizations that track this measure for internal quality improvement purposes have seen an increase in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

Numerator Statement: Ischemic stroke patients prescribed statin medication at hospital discharge
Denominator Statement: Ischemic stroke patients
Denominator Exclusions:
- Less than 18 years of age
- Length of Stay > 120 days
- Comfort measures only documented
- Enrolled in clinical trials related to stroke
- Admitted for elective carotid intervention
- Discharged to another hospital
- Left against medical advice
- Expired
- Discharged to home for hospice care
- Discharged to a health care facility for hospice care
- Documented reason for not prescribing statin medication at discharge

Measure Type: Process
Data Source: Electronic Clinical Data, Paper Medical Records
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2008 **Most Recent Endorsement Date:** Nov 01, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary or **Summary of prior review in 2012**

- The body of evidence provided supports that statin therapy should be prescribed for stroke prevention in patients with a prior history of ischemic stroke and transient ischemic attack . Based on these findings, there is a correlation between the use of statins in lowering LDL cholesterol and stoke occurrence. These findings are consistent across several clinical trials.
- 2010 AHA/ASA Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack, page 233.
    - Class I, Level of Evidence B – Statin therapy with intensive lipid-lowering effects is recommended to reduce the risk of stroke and cardiovascular events among patients with ischemic stroke or TIA who have evidence of arthrosclerosis, and LDL-c level > 100 mg/dL, and who are without known CHD.
- The 2012 Committee expressed no concerns regarding the evidence underlying this measure.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**
☐ **The developer provided updated evidence for this measure:**

Exception to evidence

N/A

Guidance from the Evidence Algorithm

Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: moderate; Consistency: high (Box 5b) → Moderate

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and voting on Evidence ?*

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following national trend data:

	CY 2010	CY2011	CY 2012	CY 2013	CY 2014
# hospitals	137	157	157	262	1296
# cases (den)	12,042	18,608	18,998	29,350	140,296
National aggregate rate	0.93	0.94	0.96	0.97	0.97
Mean hospital rate	0.89	0.90	0.94	0.95	0.96
50th percentile	0.94	0.94	0.97	0.98	0.99
10th and 90th percentiles	0.71 (10th) 1.00 (90th)	0.79 (10th) 1.00 (90th)	0.84 (10th) 1.00 (90th)	0.87 (10th) 1.00 (90th)	0.90 (10th) 1.00 (90th)

- Participation in this measure has grown significantly in the past five years. National hospital performance seems to be consistently high with little to no room for improvement.
- In 2012, the Committee noted the overall high rate of performance for this measure among reporting hospitals.

Disparities

- The developer does not provide disparities data from use of this measure.
- Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age."

Questions for the Committee:

- *Data for 2014 includes 1,296 hospitals. What proportion of hospitals caring for stroke patients is captured in this data?*
- *Performance over five years has been consistently high. How much further improvement in performance is likely using this measure?*
- *Does a gap in care still exist that warrants this national performance measure?*
- *How can this measure be used to better understand disparities in care and outcomes for certain groups?*

Preliminary rating for opportunity for improvement: ☐ High ☐ Moderate ☒ Low ☐ Insufficient

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **This is a process measure that relates directly to the desired outcome.

Maintenance - Original Endorsement Date: Jul 31, 2008 Most Recent Endorsement Date: Nov 01, 2012 - I am not aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission

**No new studies/information

1b. Performance Gap

Comments: **performance data on the measure was provided

that seems to demonstrate a gap in care and has demonstrated improvement over time which should warrant a national performance measure
data on the measure by population subgroups was not provided?
it does not demonstrate disparities in the care other than clinically but attaches a summary of data that favorably addresses the issue of disparities
**High performance since 2012; no disparities data provided
1c. High Priority (previously referred to as High Impact)
Comments: **All are stated and logical
**NA

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications
Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): medical record abstraction (paper or electronic). This is not an eMeasure.

Specifications:

- One data element defines the numerator and nine data elements define the denominator.
- Developer reports that "Since the last endorsement date, the measure specifications have been revised to align with current guideline recommendations from the American College of Cardiology (ACC)/ American Heart Association (AHA) for the management of blood cholesterol in adults. The 2013 guidelines recognize four statin benefit groups. Patients with ischemic stroke are included in the first statin benefit group along with other patients who have clinical atherosclerotic cardiovascular disease (ASCVD). For these patients, age 75 years or younger, high-intensity statin therapy is recommended unless contraindicated or patient characteristics predispose to adverse effects. The rationale for the measure was revised to reflect these recommendations, emphasizing that statin therapy is indicated for those ischemic stroke patients with evidence of atherosclerosis, i.e., 'individuals with ischemic stroke due to large artery atherosclerosis, individuals with ischemic stroke due to intrinsic small vessel disease, and individuals with ischemic stroke not directly due to atherosclerosis but with clinically evident atherosclerotic disease in an uninvolved cerebral or noncerebral bed'. The denominator statement was revised to ischemic stroke patients, and the denominator inclusions for patients with LDL greater than or equal to 100 mg/dL, or LDL not measured, or who were on a lipid-lowering medication prior to hospital arrival removed, since they are no longer relevant to current guideline recommendations. Additionally, the data elements that captured these three denominator inclusions were removed from the measure algorithm. Appendix C, Table 1.6 Lipid-Lowering Medications used for abstraction of the data element Pre-Arrival Lipid-Lowering Agent was also removed."
- Diagnosis and procedure codes have been converted from ICD-9 to ICD-10-CM diagnosis and ICD-10-PCS procedure codes. See attached spreadsheet Appendix A Table 8.1 (ischemic stroke).
- A conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications:
    - "The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015."
- Sampling monthly or quarterly is allowed; instructions for sampling are provided.
- A web-based data collection tool has been developed but is not described.

- The measure is specified at the hospital level of analysis.

Questions for the Committee :

- o *Are all the data elements clearly defined? Are all appropriate codes included?*
- o *Is the logic or calculation algorithm clear?*
- o *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

Reliability testing from prior review in 2012:

- o Inter rater-reliability of the data elements for one year (4Q2010-3Q2011) in 77 hospitals and 739 patient records showed an overall agreement rate of 96.7%.
  - o Two data elements found less than 94% agreement: Reason for Not Prescribing Statin Medication at Discharge (92%) and Pre-Arrival Lipid-Lowering Agent (92%). No Kappa scores were presented.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**
Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing see above
Results of reliability testing see above

Questions for the Committee:

- o *Developer indicates a re-test in 2012, based on an n=33, they indicated agreement of 100% when the measure focus shifted to statin medications – would additional information on that test impact your determination of the continued reliability of this measure?*
- o *The developer has not provided any new reliability testing. Does the Committee agree there is no need for repeat discussion and voting on Reliability?*

Guidance from the Reliability Algorithm
Precise specifications (Box 1) →empirical testing as specified (Box 2) → reliability testing with patient-level data elements (Box 8) → assessed all critical data elements (Box 9)→ high/moderate certainty the data elements are reliable (Box 10) → Moderate (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:

- o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

Validity testing from the prior review in 2012:

- Face validity of the data was assessed via survey and focus groups of hospitals participating in the pilot test.
- The developer reports that "All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant."
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.
- Ongoing feedback from measure users is systematically reviewed to identify trends and/or measure specifications that require clarification or revision. Predominantly, questions involving the data element Reason for Not Prescribing Statin Medication at Discharge, specifically reasons regarding the absence of documented linkage with statins, e.g., reduction in LDL-c attained with a pre-arrival lipid-lowering agent other than a statin medication.

Describe any updates to validity testing – New empirical validity testing data is presented

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- Pearson Correlation Coefficients were calculated for seven stroke measures for hospitals that reported 12 months of data and had 30 or more denominator cases for the year (2014-2015).Presumably high performing hospitals would do well on all stroke performance measures.

Validity testing results:

- Analysis of 1,318 hospitals and 2,206,379 patients records generated a table of Pearson Correlation Coefficient results that shows a statistically significant (P<.0001), positive correlation of this measure with six other stroke performance measures supporting the hypothesis that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients are excluded from the measure for the following reasons:
  - Patients less than 18 years of age
  - Patient who have a Length of Stay greater than 120 days
  - Patients with Comfort Measures
  - Patient enrolled in a Clinical Trial
  - Patients admitted for Elective Carotid Intervention
  - Patients with a documented Reason For Not Prescribing Statin Medication at Discharge
  - Patients discharged to another hospital

- Patients who left against medical advice
- Patients who expired
- Patients discharged to home for hospice care
- Patients discharged to a health care facility for hospice care

- New data on the frequency of exclusions is presented for 2, 206,379 admissions in 1,318 hospitals:
  - Comfort Measures Only: 10% (range 0.30-35.7%)
  - Clinical Trial: 0.52% (range 0.08- 10%)
  - Patients admitted for Elective Carotid Intervention: 11.5% (range 0.16-95.2%)
  - Patients with a documented Reason For Not Prescribing Statin Medication at Discharge: 3.31% (range 0.24%-29.7%)
  - Patients discharged to another hospital: 35.4% (range 0.79-76.2%)
  - Patients who left against medical advice: 0.86% (range 0.06-9.67%)
  - Patients who expired: 5.08% (range 0.39-20.1%)
  - Patients discharged to home for hospice care: 51.9% (range 6.25-94.3%)
  - Patients discharged to a health care facility for hospice care: 3% (range 0.16-23%)

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Performance data is presented above under opportunity for improvement. Complete details of the data is presented in 1b.
- The developer explains their approach: "There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations."

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods: Not applicable

2b7. Missing Data

- The developer describes the handling of missing data in the measure specifications (S.22) but does not provide information on the frequency of missing data or potential impact on measure results.

Guidance from the Validity Algorithm
Specifications consistent with evidence (Box 1) → potential threats to validity mostly assessed (Box2) →empirical validity testing (Box 3) → measure score testing (Box 6) → sound conceptualization and hypotheses (Box 7) →high confidence that scores are a valid indicator of quality (Box 8a) → High

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications
Comments: **The specifications appear consistent
**Specifications consistent with evidence
2a2. Reliability Testing
Comments: **The validity testing had adequate scope and yhe validity testing results demonstrate sufficient validity such that conclusions about quality can be made. I agree that, for most persons who have suffered an ischemic stroke, the provision of a statin is an essential element of quality
**yes adequate validity testing
2b2. Validity Testing
Comments: **All of the exclusions seem consistent with the evidence
There was no risk adjustment a stratification for this measure.
Reported analyses indicate this measure identifies meaningful differences about quality and have shown steady improvement over time
There does not appear to be missing data which would constitute a threat to the validity of this measure
**No
2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments: **The reliability testing had adequate scope and the data supports that this measure reports reliably
**Sufficient reliability

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is medical record abstraction: not all hospitals are able to generate the data electronically so a paper-based option is available
- Some data elements are in electronic form
- Data collection burden is significant.

Questions for the Committee:
- o *Are the required data elements routinely generated and used during care delivery?*
- o *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- o *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments: **All of the required data elements seem to be routinely generated and used during care delivery and, for those providers who have an EHR, are available in electronic form. The only concern about data collection is the accessibility of an EHR (which is addressed by a companion measure)
**Not all data elements are in EHR

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use

or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details :

- The Joint Commission Quality Check®: Public website that allows consumers to search for accredited and certified organizations.
- CMS Hospital Compare: Public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.
- CDC Paul Coverdell National Acute Stroke Registry: National registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.
- CMS Hospital Inpatient Quality Reporting Program (IQR): The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- The Joint Commission Annual Report-Improving America's Hospitals: Annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
- The Joint Commission Disease-Specific Care Certification for Comprehensive Stroke Centers: Certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
- The Joint Commission Disease-Specific Care Certification for Primary Stroke Centers: Certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results :

- The developer provided the following progress information on 1,296 hospitals nationwide and 140,296 patients:
    - The percentage of ischemic stroke patients with LDL > 100 mg/dL who are discharged on a statin medication has steadily increased over the past five years based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 98.7%. A modest gap of approximately 5% still exists for the 10th percentile of hospitals. This trend is consistent with increases in the rate for discharged on a statin medication published by GWTG-Stroke and PCNASR.
    - Significant and important differences persist between races in rates for patients prescribed statin therapy at discharge (Schwamm LH, et al., 2010, Qian F,et al., 2013).

Unexpected findings (positive or negative) during implementation:

- Developer did not identify unexpected findings during implementation.

Potential harms:

- The developer did not identify any unintended consequences related to this measure

Feedback :

- The developer states that feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As previously noted, measure users have indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments: **Accountability entities include 3300 joint commission accredited hospitals. The performance results are being used to further the goal of high-quality, efficient healthcare through public reporting to providers, payors and consumers.
I do not note any actual unintended consequences other than the potential that individuals who do not clinically require a statin might be placed on it to fulfill his requirement. However, for the vast majority of persons who have suffered an ischemic stroke, assuring access to a statin far outweighs the much smaller number of persons who will not benefit from the statin but for whom the statin will do no harm.
**Measure publicly reported and used in several accountability programs; high usability

Criterion 5: Related and Competing Measures

Related or competing measures

- 0118: Anti-Lipid Treatment Discharge
- 0068 : Ischemic Vascular Disease (IVD): Use of Aspirin or Another Antiplatelet
- 0074: Chronic Stable Coronary Artery Disease: Lipid Control
- 0325 : Stroke and Stroke Rehabilitation: Discharged on Antithrombotic Therapy
- 0438 : STK 05: Antithrombotic Therapy By End of Hospital Day Two
- 0543: Adherence to Statin Therapy for Individuals with Cardiovascular Disease
- 0547: Diabetes and Medication Possession Ratio for Statin Therapy
- 0639: Statin Prescribed at Discharge

Harmonization

The developer reports that the following measures address target diagnoses other than ischemic stroke or specific surgical procedures for patients 18 years or older:

- 0074 Coronary Artery Disease - provider-level measures in the ambulatory care setting.
- 0118 isolated Coronary Artery Bypass Graft (CABG) - provider-level measures in the ambulatory care setting.
- 1519 Lower Extremity Bypass (LEB) - addresses inpatient organizational performance

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0439 NQF Project: Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The focus of the measure is to reduce patient morbidity and mortality through modification of risk factors for stroke, specifically through the use of statin medications to reduce low-density lipoprotein cholesterol (LDL-c). Statin prescribed at discharge >> decreased LDL-c >> secondary stroke prevention >>decreased morbidity and mortality.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
This measure is consistent with the guidelines recommended by the American Heart Association/American Stroke Association for the prevention of stroke in patients with stroke and transient ischemic attack. An elevated serum lipid level has been a well-documented risk factor for coronary artery disease (CAD). Recently, there has been an increased focus on examining the relationship between elevated lipid levels and the incidence of stroke. In particular, some recent clinical trials have analyzed the association between lipids and non-hemorrhagic stroke. The reduction of LDL cholesterol, through lifestyle modification and drug therapy, for the prevention of strokes and other vascular events is recommended for patients with CAD in the National Cholesterol Education Program Adult Treatment Panel III (NCEP ATP III) Guidelines. In addition, recent evidence from the Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) trial supports the use of statins to lower LDL cholesterol in stroke patients without prior CAD and a fasting LDL >/= 100 mg/dL.

Based on these guidelines, all patients with ischemic stroke should have lipid profile measurement performed within 48 hours of admission unless outpatient results are available from within the past 30 days. Treatment for secondary prevention should be initiated in patients who meet NCEP ATP III criteria in the presence of LDL >/= 100 mg/dL, or continued for patients who were previously on lipid-lowering therapy and have an LDL< 100 mg/dL.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: The number of studies supporting the body of evidence for statin therapy is in the thousands. A recent editorial by Lori Mosca, MD, MPH, PhD published in Journal of the American College of Cardiology (2012) noted more than 2,300 potential studies in the literature related to sex-specific outcomes and the effects of statins on the prevention of cardiovascular disease.

The Third Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III, or ATP III) extensively analyzed the results of 63 clinical trials. Findings from these studies strongly influenced the development of ATP III recommendations (2002).

To date, the single most important study of statin therapy and stroke incidence is The Stroke Prevention by Aggressive Reduction of

Cholesterol Levels (SPARCL) trial. SPARCL randomly assigned 4731 patients who had had a stroke or TIA within one to six months before study entry, had low-density lipoprotein (LDL) cholesterol levels of 100 mg to 190 mg per deciliter (2.6 to 4.9 mmol per liter) and had no known history of coronary heart disease to double-blind treatment with 80 mg of atorvastatin per day or placebo. The incidence of fatal or non-fatal stroke in patients receiving atorvastatin was reduced by 16%. SPARCL is unique in that it is the only trial to study a cohort of patients with no known history of coronary heart disease.

A Cochrane Database of Systematic Review (10 August 2011) identified eight randomized control trials (RCTs) involving 625 participants. The review included all RCTs comparing statins (any type and dosage) versus placebo or no treatment, administered within two weeks of the onset of acute ischemic stroke or TIA. The search strategy included the Cochrane Stroke Group's Trials Register (November 2010); the Cochrane Central Register of Controlled Trials (CENTRAL)(The Cochrane Library 2010, Issue 4): MEDLINE (1950 to November 2010); and EMBASE (1980 to November 2010). In addition, ongoing trials and research registers (November 2010) were also searched and reference lists from relevant articles and contacted authors checked to further identify published, unpublished, and ongoing trials. Two review authors independently selected studies for inclusion and extracted data.
A MEDLINE search, using PubMed, of all literature related to "Hydroxymethylglutaryl-CoA Reductase Inhibitors/therapeutic use" was done for the years 2007 to 2012 and yielded 240 results. A meta-analysis of 12 studies (Biffi, et al., 2011), comprising 2013 statin users and 9682 non-users, investigated the association between prestroke statin use and clinical outcome. A meta-analysis of 18 RCTs (Kostis WJ, et al., 2012) of statins with sex-specific outcomes (N=141,235 participants, 40,275 2omen, 21,468 cardiovascular events, i.e., coronary heart disease and stroke) demonstrated the benefit of statins in decreasing morbid and mortal cardiovascular events in apparently healthy individual and in those with clinically evident cardiovascular disease.

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting the effectiveness of statin therapy on stroke outcome is moderate. There is evidence that stain therapy in both primary and secondary prevention significantly reduces subsequent major coronary events but only marginally reduces the risk of stroke recurrence. There is no clear evidence of beneficial effect from statins in those with previous hemorrhagic stroke, and it is unclear if statins should be started immediately post stroke or later (Feher, et al., 2010).

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of evidence consistently supports that stroke and TIA patients should be prescribed statin therapy for stroke prevention. No position against the importance of statin therapy for secondary stroke prevention was identified in the literature.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
Statins play an important role in brain ischemia. These drugs reduce cholesterol levels, which have been related to a reduction in vascular event risk, and also have other functions besides cholesterol metabolism, called pleiotropic effects. Statins play an important role during the acute phase of ischemia, and might have neuroprotective effects, as they act in several mechanisms during the acute phase of stroke, such as nitric oxide (NO) and glutamate metabolism, inflammation, platelet aggregation, immune responses, and apoptosis. They also have other functions which can be related, with better long-term outcome, to neurorepair mechanisms. Statins promote angiogenesis, endogenous cell proliferation, neurogenesis and new synapse formation (Rodriguez-Yáñez M, et al., 2008).

The potential positive action of stains during acute cerebrovascular ischemic event are two-fold: a neuroprotective effect, limiting damage and improving recovery; and a preventative effect on early recurrence. The SPARCL trial did report a slight increase in hemorrhagic stroke (n=55) for the atorvastatin group compared to placebo (n=33). Based on a Cochrane Review of eight RCTs of statins, no patients died from ischemic stroke or from adverse drug effects, bleeding or infections among 444 participants in six studies where these outcomes were reported. Statin treatment did not reduce all-cause mortality compared with placebo or no treatment (OR 1.51, 95% CI 0.60 60 3.81) in the 431 patients enrolled in seven studies. No cases of rhabdomyolysis occurred in 274 patients enrolled in three studies.

Diet, exercise, and lifestyle modification are assumed to be cost-effective stroke prevention strategies; however, no specific studies on the cost-effectiveness of statin therapy for the prevention of stroke recurrence were noted in the literature. Most of the literature focuses on the primary prevention of cardiovascular disease. According to Chan and colleagues (2007) recent clinical trials have found that high-dose statin therapy, compared with conventional-dose statin therapy, reduces the risk of cardiovascular events in patients with acute coronary syndromes (ACS) and stable coronary artery disease (CAD). However, the actual benefit and cost-effectiveness of

high-dose statin therapy are unknown. The daily cost difference between a high- and conventional-dose statin would need to be <$1.70, $2.65, and $3.55 to yield incremental cost-effective ratios below $50,000, $100,000, and $150,000 per quality-adjusted life year (QALY).

Another study (Lazar LD, et al., 2011), evaluated the cost-effectiveness of statin therapy for primary prevention. This study utilized the Coronary Heart Disease (CHD) Policy Model, an established computer simulation, Markov state-transition model of CHD incidence, prevalence, mortality and costs in the US population >35 years of age. Cost-savings were projected for persons with LDL > 100 mg/dL and moderately high risk, LDL > 130 mg/dL and moderate risk, and LDL > 160 mg/dL and lower or lowest risk with the cost of statins at $4/month. Using this strategy, it would be possible to prevent 14,000 CHD deaths per year and save over $1.4 billion a year compared with current levels of treatment.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: National Cholesterol Education Program, Adult Treatment Panel (ATP) III (2002). During our systematic review, it was determined that the guideline developers accounted for a balanced representation of information, and provided information that was accessible and met the requirements set out in this measure maintenance form.

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: The Third Report of the Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III, or ATP III) presents the National Cholesterol Education Program's (NCEP's) updated recommendations for cholesterol testing and management.
ATP III contains both evidence statements and specific recommendations based on these statements. Each evidence statement is qualified according to category of evidence (A-D) and strength of evidence (1-3), as follows:

Type of Evidence
Category and Description of Type of Evidence
A – Major randomized controlled clinical trials (RCTs)
B – Smaller RCTs and meta-analyses of other clinical trials
C – Observational and metabolic studies
D – Clinical experience

Strength of Evidence
Category and Description of Strength of Evidence
1 – Very strong evidence
2 – Moderately strong evidence
3 – Strong trend

1c.13 Grade Assigned to the Body of Evidence: A1 (Major randomized controlled clinical trials (RCTs)/Very strong evidence) // B1 (Smaller RCTs and meta-analyses of other clinical trials/Very Strong evidence)

1c.14 Summary of Controversy/Contradictory Evidence: There is some recent debate about the use of statins in women who do not have heart disease but do have high levels of cholesterol. An observational study (Culver AL, et al., 2012), involving 153,840 postmenopausal women enrolled in the Women's Health Initiative, reported a 48% greater risk of developing diabetes mellitus for women taking statin medications when compared to women not taking the medication at baseline. However, a new meta-analysis (Kostis WJ, et al., 2012) of 18 trials with 141,235 participants concluded that statin therapy was beneficial in both the primary and secondary prevention settings for both men and women.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- Adams RJ, Albers G, Alberts MJ, Benavente O, Furie K, Goldstein LB, Gorelick P, Halperin J, Harbaugh R, Johnston SC, Katzan I, Kelly-Hayes M, Kenton EJ, Marks M, Sacco RL, Schwamm LH. Update to the AHA/ASA recommendations for the prevention of stroke in patients with stroke and transient ischemic attack. Stroke. 2008;39(5):1650.
- Biffi A, Devan WJ, Anderson CD, Cortellini L, Furie KL, Rosand J, Rost NS. Statin treatment and functional outcome after ischemic stroke: case-control and meta-analysis. Stroke. 2011;42(5): 1314-9.

- Chan PS, Nallamothu BK, Gurm HS, Hayward RA, Vijan S. Incremental benefit and cost-effectiveness of high-dose statin therapy in high-risk patients with coronary artery disease. Circulation. 2007;115:2398-2409.
- Craig SR, Amin RV, Russell DW, Paradise NF. Blood cholesterol screening influence of fasting state on cholesterol results and management decisions. J Gen Intern Med. 2000 Jun;15(6):395-9.
- Culver AL, Ockene IS, Balasubramanian R, Olendzki BC, Sepavich DM, Wactawski-Wende J, Manson JE, Qiao Y, Liu S, Merriam PA, Rahilly-Tierny C, Thomas F, Berger JS, Ockene JK, Curb JD, Ma Y. Statin Use and Risk of Diabetes Mellitus in Postmenopausal Women in the Women´s Health Initiative.
Arch Intern Med. 2012; 172: 144 – 152.
- Feher A, Pusch G, Koltai K, Tibold A, Gasztonyi B, Szapary L, Feher G. Statintherapy in the primary and secondary prevention of ischaemic cerebrovascular diseases. Int J Cardiol. 2011; 148(2): 131-8.
- Feinberg WM, Albers GW, Barnett HJM, et al. Guidelines for the Management of Transient Ischemic Attacks. From the Ad Hoc Committee on Guidelines for the Management of Transient Ischemic Attacks of the Stroke Council of the American Heart Association. 1994.
- Gore JM, Goldberg RJ, Matsumoto AS, et al. Validity of serum total cholesterol level obtained within 24 hours of acute myocardial infarction. Am J Cardiol. 1984;54:722-725.
- Grundy SM, Cleeman JI, Merz CNB, Brewer, HB, et. al. Implications of Recent Clinical Trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. Circulation. 2004;110:227-239
- High-Dose Atorvastatin after Stroke or Transient Ischemic Attack. (New England Journal of Medicine. NEJM Vol. 355 2006:549-559.
- Kostis WJ, Cheng JQ, Dobrzynski BA, Cabrera J, Kostis JB. Meta-analysis of statin effects in women versus men. Journal of American College of Cardiology. 2012;59(6):572-82.
- Lazar LD, Pletcher MJ, Coxson PG, Bibbins-Domingo K, Goldman L. Cost-effectiveness of statin therapy for primary prevention in a low-cost statin era. Circulation. 200;124:146-153.
- Mitka M. Some queston use of statins to reduce cardiovascular risks in healthy women. JAMA. 2012;37(9):893-94.
- Pitt B, Loscalzo J, Ycas J, Raichlen JS. Lipid Levels After Acute Coronary Syndromes. J Am Coll Cardiol 2008;51;1440-1445.
- Rodrigues-Yáñez M, Agulla J, Rodrigues-González R, Sobrino, T, Castillo J. Statins and stroke. Ther Adv Cardiovasc Dis. 2008;2(3): 157-66.
- Sacco RL, Adams R, Albers G, Alberts MJ, Benavente O, Furie K, Goldstein LB, Gorelick P, Halperin J, Harbaugh R, Johnston SC, Katzan I, Kelly-Hayes M, Kenton EJ, Marks M, Schwamm LH, Tomsick T. Guidelines for Prevention of Stroke in Patients With Ischemic Stroke or Transient Ischemic Attack: A Statement for Healthcare Professionals From the American Heart Association/American Stroke Association Council on Stroke: Co-Sponsored by the Council on Cardiovascular Radiology and Intervention. Stroke. Vol. 37, 2006:577.
- Schellinger PD, Bryan RN, Caplan LR, Detre JA, et al. Evidence-based guideline: The role of diffusion and perfusion MRI for the diagnosis of acute ischemic stroke: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. Neurology 2010;75;177-185.
- Squizzato A, Romualdi E, Dentali F, Ageno W. Statins for acute ischemic stroke. Cochrane Database Syst Rev. 2011 Aug 10;(8): CD007551.
- Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Final Report Circulation Vol. 106 2002: 3143-3421.
- Van Dis FJ, Keilson LM, Rundell CA, et al. Direct measurement of serum low-density lipoprotein cholesterol in patients with acute myocardial infarction on admission to the emergency room. Am J Cardiol. 1996;77:1232-1234.
- Weiss R, Harder M, Rowe J. The relationship between nonfasting and fasting lipid measurements in patients with or without type 2 diabetes mellitus receiving treatment with 3-hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors. Clin Ther. 2003 May;25(5):1490-7.

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)*:
2010 Update to the AHA/ASA Recommendations for the Prevention of Stroke in Patients with Stroke and Transient Ischemic Attack, page 233.
Class I, Level B Recommendation
Statin therapy with intensive lipid-lowering effects is recommended to reduce the risk of stroke and cardiovascular events among patients with ischemic stroke or TIA who have evidence of atherosclerosis, and LDL-c level > 100 mg/dL, and who are without known CHD.

1c.17 Clinical Practice Guideline Citation: Furie KL, Kasner SE, Adams RJ, Albers GW, et al. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack. A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2011;42:232-33.

1c.18 National Guideline Clearinghouse or other URL: http://stroke.ahajournals.org/content/42/1/227.full.pdf

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: This guideline was approved by the American Heart Association Science Advisory and Coordinating Committee on July 13, 2010. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: American Heart Association (AHA)/American Stroke Association (ASA)recommendation rating scale:

LEVEL A – Multiple populations evaluated.* Data derived from multiple randomized clinical trials or meta-analysis.

LEVEL B – Limited populations evaluated.* Data derived from a single randomized trial or nonrandomized studies.

LEVEL C – Very limited populations evaluated.* Only consensus opinion of experts, case studies, or standard of care.

CLASS I Benefit >>> Risk Procedure/Treatment SHOULD be performed/administered

CLASS IIa Benefit >> Risk Additional studies with focused objectives needed IT IS REASONABLE to perform procedure/administer treatment

CLASS IIb Benefit > Risk Additional studies with broad objectives needed; additional registry data would be helpful Procedure/Treatment MAY BE CONSIDERED

CLASS III Risk > Benefit Procedure/Treatment should NOT be performed/administered SINCE IT IS NOT HELPFUL AND MAY BE HARMFUL

CLASS I, LEVEL A – Recommendation that procedure or treatment is useful/effective. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS IIa, LEVEL A – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from multiple randomized trials or meta-analysis.

CLASS IIb, LEVEL A – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from multiple randomized trials or meta-analysis.

CLASS III, LEVEL A – Recommendation that procedure or treatment is not useful/effective and may be harmful. Sufficient evidence from multiple randomized trials or meta-analysis.

CLASS I, LEVEL B – Recommendation that procedure or treatment is useful/effective. Evidence from single randomized trial or nonrandomized studies.

CLASS IIa – LEVEL B – Recommendation in favor of treatment or procedure being useful/effective. Some conflicting evidence from single randomized trial or nonrandomized studies.

CLASS IIb, LEVEL B – Recommendation's usefulness/efficacy less well established. Greater conflicting evidence from single randomized trial or nonrandomized studies.

CLASS III, LEVEL B – Recommendation that procedure or treatment is not useful/effective and may be harmful. Evidence from single randomized trial or nonrandomized studies.

CLASS I, LEVEL C – Recommendation that procedure or treatment is useful/effective. Only expert opinion, case studies, or standard of care.

CLASS IIa – LEVEL C – Recommendation in favor of treatment or procedure being useful/effective. Only diverging expert opinion, case studies, or standard of care.

CLASS IIb, LEVEL C – Recommendation's usefulness/efficacy less well established. Only diverging expert opinion, case studies, or standard of care.

CLASS III, LEVEL C – Recommendation that procedure or treatment is not useful/effective and may be harmful. Only expert opinion, case studies, or standard of care.

1c.23 Grade Assigned to the Recommendation: Class I, Level of Evidence A/B

1c.24 Rationale for Using this Guideline Over Others: The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: High 1c.26 Quality: Moderate1c.27 Consistency: High

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0439_Evidence_MSF5_0_Data.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fourth leading cause of death in the United States and a leading cause of serious, long-term disability, associated with significant costs. Recent guidelines recommend high-intensity statin therapy for patients with ischemic stroke unless contraindicated or patient characteristics require dosage modification.

Healthcare organizations that track this measure for internal quality improvement purposes have seen an increase in the measure rate over time. This measure has been included in the CMS Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
In October, 2009, The Joint Commission added the stroke (STK) measure set as a new core measure option to meet performance measure requirements for Joint Commission hospital accreditation purposes. Below is the specified level of analysis for STK-6 beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 979 denominator cases; 876 numerator cases; 48 hospitals; 0.89479 national aggregate rate; 0.88567 mean of hospital rates; 0.18656 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.94949 50th percentile rate/median rate; 0.85165 25th percentile rate/lower quartile; and, 0.66667 10th percentile rate.

CY 2010: 12,042 denominator cases; 11,188 numerator cases; 137 hospitals; 0.92908 national aggregate rate; 0.89366 mean of hospital rates; 0.12881 standard deviation; 1.0 90th percentile rate; 0.97059 75th percentile rate/upper quartile; 0.94231 50th percentile rate/median rate; 0.86667 25th percentile rate/lower quartile; and, 0.70647 10th percentile rate.

CY 2011: 18,608 denominator cases; 17,512 numerator cases; 157 hospitals; 0.9411 national aggregate rate; 0.90293 mean of hospital rates; 0.12829 standard deviation; 1.0 90th percentile rate; 0.97297 75th percentile rate/upper quartile; 0.94361 50th percentile rate/median rate; 0.88889 25th percentile rate/lower quartile; and, 0.78571 10th percentile rate.

CY 2012: 18,998 denominator cases; 18,165 numerator cases; 157 hospitals; 0.95615 national aggregate rate; 0.9412 mean of hospital rates; 0.08087 standard deviation; 1.0 90th percentile rate; 0.99371 75th percentile rate/upper quartile; 0.96825 50th percentile rate/median rate; 0.925 25th percentile rate/lower quartile; and, 0.84375 10th percentile rate.

CY 2013: 29,350 denominator cases; 28,374 numerator cases; 262 hospitals; 0.96675 national aggregate rate; 0.9469 mean of hospital rates; 0.09876 standard deviation; 1.0 90th percentile rate; 0.99649 75th percentile rate/upper quartile; 0.97938 50th percentile rate/median rate; 0.94444 25th percentile rate/lower quartile; and, 0.86585 10th percentile rate.

CY 2014: 140,296 denominator cases; 136,545 numerator cases; 1296 hospitals; 0.97326 national aggregate rate; 0.95761 mean of hospital rates; 0.08458 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.98507 50th percentile rate/median rate; 0.95303 25th percentile rate/lower quartile; and, 0.89474 10th percentile rate.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
Not applicable

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al. 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

In the national REasons for Geographic and Racial Differences in Stroke (REGARDS) cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038): and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years of age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age : RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer,et al., 2010).

In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).

Differences in access to and use of stroke prevention therapies in all racial and ethnic groups has been poorly documented and understudied; however, one study of 5840 stroke survivors as part of the National Health Interview Survey, found that women, blacks or African Americans, and the poor were significantly less likely to fill prescriptions because of costs. Disparities are reduced when patients have health insurance and ready access to care.

According to the literature, minorities are less likely to receive medications for secondary prevention, including statin therapy for hyperlipidemia (Cruz-Flores, et al., 2011). Yood and colleagues (2006) found that blacks or African Americans newly diagnosed with dyslipidemia and prescribed statins were 36% less likely to achieve low-density lipoprotein goals over time (hazard ratio 0.64, 95% CI O.61 to 0.68). This disparity persisted after low-density lipoprotein testing and adjustment for statin adherence (hazard ratio 0.60, 95% CI 0.57 to 0.63). Another study by Mark and associates (2007) noted that blacks or African Americans were less likely than whites to be switched between lipid-lowering agents (OR 0.68 95% CI 0.60 to 0.78), to have treatment adjusted (OR 0.53, 95% CI 0.43 to 0.66), or to be prescribed higher medication dosages (OR 0.75, 95% CI 0.67 to 0.84).

Since the last endorsement date, Schwamm and colleagues (2010) found that black patients with stroke received fewer evidence-based care processes than Hispanic or white patients. Lipid-lowering therapy at discharge for patients with low-density lipoprotein (LDL) > 100, or those on lipid-lowering agents before hospital admission, or in whom LDL was not measured in the past 30 days, was one care process evaluated in this study. Blacks had lower odds relative to white patients of receiving lipid therapy at discharge after adjustment for both patient and hospital level variables: OR 0.91 [95% CI 0.88-0.96]

A more recently published study (Qian F, et al, 2013) from GWTG also noted statistically significant (P<0.01) racial and ethnic disparities for lipid-lowering therapy at discharge. Using patient data (n=200,900) from the American Heart Association/American Stroke Association Get With The Guidelines (GWTG)-Stroke program from April 2003 through December 2008, the following performance measure rates were reported for lipid-lowering medication prescribed at discharged : non-Hispanic White (n=170,694) 84.3%; non-Hispanic Black (n=20,514) 83.8%; Hispanic (n=6632) 85.8%; and non-Hispanic Asian American (n=3060) 86.7%.

2014 data from the Paul Coverdell National Acute Stroke Registry (PCNASR) reported greater disparity for women (95.9%) compared to men (97.0%) prescribed statin therapy at discharge. Similar rates of statin therapy prescribed at discharge were reported for white patients (96.3%) and other races (96.7%); aggregate rate 96.4%; n=45,249 (Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, 2014).

• Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, Yan T, Escarce JJ. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-8.
• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, Griffith P, Gorelick PB, Howard G, Leira EC, Morgenstern LB, Ovbiagele B, Peterson E, Rosamond W, Trimble B, Valderrama AL, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• Gillum RF,Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 2011;42(11):3294-6.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-75.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Karve S, Balkrishnan R, Seiber E Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle

Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. Stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Mark TL, Axelsen KJ, Mucha L, Sadkova Y. Racial differences in switching, augmentation, and titration of lipid-lowering agents by Medicare/Medicaid dual-eligible patients. Am J Manag Care. 2007;13(3):S72-S79.
• Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, Kerber KA, Skolarus LE, Meurer WJ, Burke JF, Adelman EE, Baek J, Lisabeth LD. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol. 2013;74:778–785.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Schwamm LH, Syed FA, Reeves MJ, Smith EE, Saver JL, Messe S, Bhatt DL, Grau-Sepulveda MV, Peterson ED, Fonarow GC. Temporal trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With the Guidelines-Stroke hospitals. Circ Cardiovasc Quality Outcomes. 2013;6:543-549.
• Yood MU, McCarthy BD, Kempf J, Kucera GP, Wells K, Oliveria S, Stang P. Racial differences in reaching target low-density lipoprotein goal among individuals treated with statin therapy. Am Heart J. 2006;152:777-784.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality, Severity of illness

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States remains attributable to stroke. More women than men die of stroke each year. Women accounted for almost 60% of US stroke deaths in 2008 (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, et al., 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 ($P<0.05$). (US Burden of Disease Collaborators, 2013) . Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

According to the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III/ATP III) (National Institutes of Health, 2002), statins should be considered as first-line drugs when LDL-lowering drugs are indicated to achieve LDL treatment goals. HMG CoA reductase inhibitors (statins) are powerful LDL-lowering drugs. Statin therapy reduces the risk of acute coronary syndromes, coronary procedures, and other coronary outcomes in both primary and secondary prevention. It also reduces the risk of stroke in secondary prevention.

The Stroke Prevention by Aggressive Reduction of Cholesterol Levels (SPARCL) trial (Amarenco, P, et al., 2006) concluded that in patients with recent stroke or TIA and without known coronary heart disease, 80 mg of atorvastatin per day reduced the overall incidence of strokes and cardiovascular events, despite a small incidence of hemorrhagic stroke. The trial randomly assigned 4731 patients who had had a stroke or TIA within one to six months before study entry, had low-density lipoprotein (LDL) cholesterol levels of 100 to 190 mg per deciliter (2.6 to 4.9 mmol per liter), and had no known coronary heart disease to double-blind treatment with 80 mg of atorvastatin per day or placebo. The primary end point was a first nonfatal or fatal stroke. The mean LDL cholestrerol level during the trial was 73 mg per deciliter (1.9 mmol per liter) among patients receiving atorvastatin and 129 mg per deciliter (3.3 mmol per liter) among patients receiving placebo. During a median follow-up of 4.9 years, 265 patients (11.2 percent) receiving atorvastatin and 311 patients (13.1 percent) receiving placebo had a fatal or nonfatal stroke (5-year absolute reduction in risk, 2.2 percent; adjusted hazard ratio, 0.84; 95 percent confidence interval, 0.71 to 0.99; P=0.03; unadjusted P=0.05) The atorvastatin group had 218 ischemic strokes and 55 hemorrhagic strokes. The five-year absolute reduction in the risk of major cardiovascular events was 3.5 percent (hazard ratio, 0.80; 95 percent confidence interval, 0.69 to 0.92; P=0.002). The overall mortality rate was similar, with 216 deaths in the atorvastatin group and 211 deaths in the placebo group (P=0.98), as were the rates of serious adverse events. Elevated liver enzyme values were more common in patients taking atorvastatin.

In 2013, the American College of Cardiology (ACC) / American Heart Association (AHA) updated the Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults, and strongly recommended high-intensity statin therapy for secondary prevention in patients with clinical atherosclerotic cardiovascular disease (ASCVD), unless contraindicated. Patients with ischemic stroke due to atherosclerosis are included in this first of four statin benefit groups. High-intensity statin therapy should be initiated or continued as first-line therapy in both women and men =75 years. When high-intensity statin therapy is contraindicated or for those patients unable to tolerate high-intensity statin therapy, moderate-intensity statin therapy should be used as a second option. High-intensity therapy may also be reasonable for ischemic stroke patients > 75 years, if risk-reduction benefits outweigh the risk of adverse events and the patient can tolerate it (Stone NJ, et al, 2013). AHA/ASA Guidelines for the Prevention of Stroke in Patients with Stroke or Transient Ischemic Attack (Kernan WN, et al, 2014) were revised the following year to align with 2013 ACC/AHA recommendations. Statin therapy was recommended to reduce the risk of stroke and cardiovascular events for patients with ischemic stroke or TIA presumed to be of atherosclerotic origin and LDL > 100 with or without evidence for other ASCVD, as well as, patients with LDL < 100 and no evidence for other clinical ASCVD.

1c.4. Citations for data demonstrating high priority provided in 1a.3

• Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.

• George M, Xin T, McGruder H, Yoon P, Rosamond W, Winquist, A., Hinchey J, Wal, H, Pandey D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.

• Grundy SM, Cleeman JI, Merz CNB, Brewer, HB, et. al. Implications of Recent Clinical Trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. Circulation. 2004;110:227-239

• High-Dose Atorvastatin after Stroke or Transient Ischemic Attack. (New England Journal of Medicine. NEJM Vol. 355 2006:549-559.

• Kernan WN, Ovbiagele B, Black HR, Bravata DM, Chimowitz MI, Ezekowitz MD, Fang MC, Fisher M, Furie KL, Heck DV, Johnston SC, Kasner SE, Kittner SJ, Mitchell PH, Rich MW, Richardson D, Schwamm LH, Wilson JA. Guidelines for the Prevention of Stroke in Patients With Stroke or Transient Ischemic Attack: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. Stroke. 2014;45:241-2160-2236.

• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.

• Stone NJ, Robinson J, Lichtenstein AH, Noel Bairey Merz C, Blum CB, Eckel RH, Goldberg AC, Gordon D, Levy D, Lloyd-Jones DM, McBride P, Schwartz JS, Shero Jr, ST, Smith SC, Watson K, Wilson PWF. "Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Disease in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. [In eng]. Circulation 11, (Nov 2013): 1-84.

• Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Final Report Circulation Vol. 106 2002: 3143-3421.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Appendix_A.1-635878758534627046.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Since the last endorsement date, the measure specifications have been revised to align with current guideline recommendations from the American College of Cardiology (ACC) / American Heart Association (AHA) for the management of blood cholesterol in adults (Stone, et al, 2013). The 2013 guidelines recognize four statin benefit groups. Patients with ischemic stroke are included in the first statin benefit group along with other patients who have clinical atherosclerotic cardiovascular disease (ASCVD). For these patients, age 75 years or younger, high-intensity statin therapy is recommended unless contraindicated or patient characteristics predispose to adverse effects. The rationale for the measure was revised to reflect these recommendations, emphasizing that statin therapy is indicated for those ischemic stroke patients with evidence of atherosclerosis, i.e., "individuals with ischemic stroke due to large artery atherosclerosis, individuals with ischemic stroke due to intrinsic small vessel disease, and individuals with ischemic stroke not directly due to atherosclerosis but with clinically evident atherosclerotic disease in an uninvolved cerebral or noncerebral bed". The denominator statement was revised to ischemic stroke patients, and the denominator inclusions for patients with LDL greater than or equal to 100 mg/dL, or LDL not measured, or who were on a lipid-lowering medication prior to hospital arrival removed, since they are no longer relevant to current guideline recommendations. Additionally, the data elements that captured these three denominator inclusions were removed from the measure algorithm. Appendix C, Table 1.6 Lipid-Lowering Medications used for abstraction of the data element Pre-Arrival Lipid-Lowering Agent was also removed.

All ICD-9-CM diagnosis codes and ICD-9-CM procedure codes were converted to ICD-10-CM diagnosis and ICD-10-PCS procedure codes throughout the measure specifications. The Department of Health and Human Services (HHS) mandated that all entities covered by the Health Insurance Portability and Accountability Act (HIPAA) must transition to a new set of codes for electronic health care transactions on October 1, 2015.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Ischemic stroke patients prescribed statin medication at hospital discharge

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
One data element is used to calculate the numerator:
• Statin Medication Prescribed at Discharge – Documentation that a statin medication was prescribed at hospital discharge. Allowable values: Yes, No/UTD or unable to determine from medical record documentation.

Patients are eligible for the numerator population when the allowable value equals "yes" for the data element.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Ischemic stroke patients

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Nine data elements are used to calculate the denominator:

1. Admission Date – The month, day and year of admission to acute inpatient care.
2. Birthdate - The month, day and year the patient was born.
3. Clinical Trial - Documentation that during this hospital stay the patient was enrolled in a clinical trial in which patients with stroke were being studied. Allowable values: Yes or No/UTD.
4. Comfort Measures Only – The earliest day the physician/APN/PA documented comfort measures only after hospital arrival. Allowable values: 1 (Day 0 or 1); 2 (Day 2 or after); 3 (Timing Unclear); 4 (Not Documented/UTD).
5. Discharge Date – The month day and year the patient was discharged from acute care, left against medical advice or expired during the stay.
6. Discharge Disposition – The place or setting to which the patient was discharged on the day of hospital discharge.
7. Elective Carotid Intervention – Documentation demonstrates that the current admission is solely for the performance of an elective carotid intervention (e.g., elective carotid endarterectomy, angioplasty, carotid stenting).
Allowable values: Yes or No/UTD.
8. ICD-10-CM Principal Diagnosis Code - The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.
9. Reason For Not Prescribing Statin Medication at Discharge – Documentation of a reason for not prescribing a statin medication at discharge.
Allowable values: Yes or No/UTD.

Population: Discharges with ICD-10-CM Principal Diagnosis Code for ischemic stroke as defined in Appendix A, Table 8.1.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
- Less than 18 years of age
- Length of Stay > 120 days
- Comfort measures only documented
- Enrolled in clinical trials related to stroke
- Admitted for elective carotid intervention
- Discharged to another hospital
- Left against medical advice
- Expired

- Discharged to home for hospice care
- Discharged to a health care facility for hospice care
- Documented reason for not prescribing statin medication at discharge

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

• The patient age in years is equal to the Discharge Date minus the Birthdate. Patients less than 18 years are excluded.
• The Length of Stay (LOS) in days is equal to the Discharge Date minus the Admission Date. If the LOS is greater than 120 days, the patient is excluded.
• Patients with Comfort Measures Only allowable value of 1 (Day 0 or 1), 2 (Day 2 or after), and 3 (Timing unclear) are excluded.
• Patients are excluded if "Yes" is selected for Clinical Trial.
• Patients with ICD-10-PCS procedure codes for carotid intervention procedures as identified in Appendix A, Table 8.3,, if medical record documentation states that the patient was admitted for the elective performance of this procedure are excluded.
• Patients with Discharge Disposition allowable value of 2 (Hospice-Home), 3 (Hospice-Health Care Facility), 4 (Acute Care Facility), 6 (Expired), or 7 (Left Against Medical Advice/AMA) are excluded.
• Patients are excluded if "Yes" is selected for Reason For Not Prescribing Statin Medication at Discharge.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*

Not applicable, the measure is not stratified.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)

No risk adjustment or risk stratification

If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*

Not applicable.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*

Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

Not applicable.

S.16. **Type of score:**

Rate/proportion

If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*

1. Start processing. Run cases that are included in the Stroke (STK) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check ICD-10-CM Principal Diagnosis Code

a. If the ICD-10-CM Principal Diagnosis Code is not on Table 8.1, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.

b. If the ICD-10-CM Principal Diagnosis Code is on Table 8.1, continue processing and proceed to Discharge Disposition.

3. Check Discharge Disposition
a. If Discharge Disposition equals 2, 3, 4, 6, 7 the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
b. If Discharge Disposition equals 1, 5, 8, continue processing and proceed to Comfort Measures Only.

4. Check Comfort Measures Only
a. If Comfort Measures Only is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Comfort Measures Only equals 1, 2, or 3, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Comfort Measures Only equals 4, continue processing and proceed to Clinical Trial.

5. Check Clinical Trial
a. If Clinical Trial is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Clinical Trial equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If Clinical Trial equals No, continue processing and proceed to Elective Carotid Intervention.

6. Check admitted for Elective Carotid Intervention
a. If Elective Carotid Intervention is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Elective Carotid Intervention equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Elective Carotid Intervention equals No, continue processing and proceed to Pre-Arrival Lipid-Lowering Agent.

7. Check Statin Medication Prescribed at Discharge
a. If Statin Medication Prescribed at Discharge is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Statin Medication Prescribed at Discharge equals Yes, the case will proceed to a Measure Category Assignment of E and will be in the Numerator Population. Stop processing.
c. If Statin Medication Prescribed at Discharge equals No, continue processing and check Reason for Not Prescribing Statin Medication at Discharge.

8. Check Reason for Not Prescribing Statin Medication at Discharge
a. If Reason for Not Prescribing Statin Medication at Discharge is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Reason for Not Prescribing Statin Medication at Discharge equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Reason for Not Prescribing Statin Medication at Discharge equals No, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter for the measure set cannot sample.

Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

Quarterly Sampling
The Quarterly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing quarterly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If "N" >/= 900, then "n" 180
If "N" 226-899, then "n" 20% of Initial Patient Population size
If "N" 45-225, then "n" 45
If "N" 6-44, No sampling; 100% Initial Patient Population required
If "N" 0-5, Submission of patient level data is not required; if submission occurs, 100% Initial Patient Population required

Monthly Sampling
The Monthly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing monthly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If "N >/= 300, then "n" 60
If "N 76-299, then "n" 20% of Initial Patient Population size
If "N 15-75, then "n" 15
If "N < 15, No sampling; 100% Initial Patient Population required

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Each data element in the data dictionary includes suggested data sources. The data are collected using contracted Performance Measurement Systems (vendors) that develop data collection tools based on the measure specifications. The tools are verified and tested by Joint Commission staff to confirm the accuracy and conformance of the data collection tool with the measure specifications. The vendor may not offer the measure set to hospitals until verification has been passed.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0439_MeasureTesting_MSF5.0_Data-635905388861973443.doc

NQF #: 0439 NQF Project: Neurology Project

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on this measure are as follows:
170 health care organizations representing various hospital demographics:
17 For Profit; 144 Not for Profit; 9 Government
62 >=300 beds; 71 100-299 beds; 37 <100 beds
142 Urban; 28 Rural
22 Teaching; 148 Non-teaching

States represented in this data collection effort include: AL, AR, AZ, CA, CT, FL, GA, IA, ID, IL, IN, KY, LA, MA, MD, MI, MN, MO, MS, NC, NE, NJ, NM, NY, OH, PA, PR, SC, SD, TN, TX, VA, WA, WI

15 performance measurement systems are used for data transmission to The Joint Commission.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with data transmission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of patient records over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rate of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify potential measure enhancements.

This measure was re-tested in 2010-2011 (n=33), following modification of the measure to focus specifically on statin medications (as opposed to the more general designation of lipid lowering medications, as had been originally tested). Retrospective comparison of vendor re-abstracted data and data submitted to The Joint Commission warehouse by the healthcare organization revealed zero defects. The agreement rate was 100% for all STK-6 Discharged on Statin Medication data elements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures). The following is a list of the major tests done on the submitted ORYX data:

- Transmission of complete data

- Usage of data received: To understand if the HCO provides the relevant service to treat the relevant population
- Investigation of aberrant data points
- Verification of patient population and sample size
- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity
- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors as outlined in the contract is as follows:

- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Data element agreement rates reported to The Joint Commission for the time period of one year (4Q2010 – 3Q2011) have shown an overall agreement rate of 96.7%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-6.

Data Elements	Total 'n' numerator	Total 'n' denominator	Rate
Clinical Trial	712	714	99.7%
Comfort Measures Only	709	714	99.3%
Elective Carotid Intervention	711	714	99.6%
LDL-c Greater Than or Equal to 100 mg/dL	363	380	95.5%
LDL-c Measured Within the First 48 Hours or 30 Days Prior to Hospital Arrival	434	463	93.7%
Pre-Arrival Lipid-Lowering Agent	426	464	91.8%
Reason for Not Prescribing Statin Medication at Discharge	91	99	92.0%
Statin Medication Prescribed at Discharge	416	441	94.3%

These agreement rates are considered to be well within acceptable levels.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* are consistent with the evidence cited in support of the measure focus (*criterion 1c)* and identify any differences from the evidence:
This measure focuses on determining the proportion of ischemic stroke patients with LDL greater than or equal to 100 mg/dL, or LDL not measured, or who were on a lipid-lowering medication prior to hospital arrival that are prescribed a statin medication at hospital discharge. The literature supports the effectiveness of statin medications for reducing serum levels of low-density lipoprotein (LDL) cholesterol circulating in the blood. Statins are a class of pharmaceutical agents that modify LDL cholesterol by blocking the action of the liver enzyme needed to synthesize cholesterol. Elevated LDL is associated with increased risk offatty plaque build-up in the walls of arteries which can result in decreased blood flow to the brain and stroke.

Statin medications are the first-line drugs of choice for reducing LDL for ischemic stroke patients with LDL levels greater than or equal to 100 mg/dL as recommended in AHA clinical practice guidelines. Randomized clinical trials (i.e., Stroke Prevention by Aggressive Reduction in Cholesterol Levels – "SPARCL" and Heart Protection Study – "HPS") support the use of statins in patients with large artery atherosclerotic or small artery branch atherosclerotic (lacunar) stroke. These trials convincingly demonstrated that intensive lipid-lowering therapy using statin medication was associated with a dramatic reduction in the rate of recurrent ischemic stroke and major coronary events. There is no published evidence to recommend the routine use of statins in the treatment of stroke patients who do not have atherosclerosis and do not otherwise qualify for lipid lowering due to other conditions.

Patients less than 18 years of age that have a length of stay (LOS) more than 120 days, enrolled in a clinical trial for stroke or who were designated "comfort measures only" anytime during hospitalization are excluded. In addition, patients who were discharged to a health care facility for hospice care, home for hospice care, who expired or who left against medical advice are excluded to harmonize with other CMS/Joint Commission measures.

Operationally, there are two differences between the measure specifications and guideline recommendations. First, patients with a planned admission to the hospital for an elective carotid intervention which restores blood flow to the brain and prevents stroke, such as carotid endarterectomy or carotid stenting, are excluded from the measure. Second, patients with a contraindication to statin therapy or a documented reason why statin therapy is not indicated are excluded from the measure.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:
At the time this measure was originally tested measure validity was assessed via survey and focus groups of hospitals participating in the pilot test. All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant.

Since the measure has been in national use, continued validity of the measure has been determined through analysis of feedback from measure users. The Joint Commission provides a web-based application with which measure users can provide feedback regarding appropriateness of measure specifications, request clarification of specifications, and/or provide other comments pertinent to the measure. This feedback is systematically, continually, reviewed in order to identify trends and to identify areas of the measure specifications that require clarification or revision. Additionally, Joint Commission staff continually monitors the national literature and environment in order to assess continued validity of this measure. An expert technical advisory panel (TAP) meets on a quarterly basis in order to assess continued validity of this measure. And finally, the conversion from ICD-9-CM diagnosis codes to ICD-10-CM diagnosis codes has been completed and reviewed by the Technical Advisory Panel for validity. The panel has determined that the intent of the measure has not changed as a result of the conversion. The crosswalk will also be posted in a future version of the specifications manual for public comment, and results of feedback will be reviewed and incorporated into the measure specifications where indicated.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:
Analysis of feedback obtained via our automated feedback system reveals 37 submissions regarding specifications for this measure over the past year. Questions regarding clarification of the data element Reason for Not Prescribing Statin Medication at Discharge are most common, specifically reasons that are inferred in the absence of documented linkage with statins, e.g., reduction in LDL-c attained with a pre-arrival lipid-lowering agent other than a statin medication.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:
Measure exclusions that were not derived directly from the evidence are presented below. Please note that these are population exclusions that are necessary to ensure consistency in all measures in this 8 measure set.

These exclusions were analyzed for frequency of occurrence. An issue that is of great concern to users of this measure is that due to the presence of exceptions to the measure, attainment of a 100% measure rate is not possible. Because of the role of this measure in

the current Joint Commission accreditation process this is especially troubling to measure users. This concern is the basis for a number of the non-evidence-based exclusions to these measures. The following measure exclusions that were not derived directly from the evidence are as follows:

1. Patients < 18 years old
2. Patients who have a length of stay (LOS) greater than 120 days
3. Patients with Comfort Measures Only documented
4. Patients enrolled in clinical trials
5. Patients admitted for Elective Carotid Intervention
6. Patients discharged to another hospital (acute care facility)
7. Patients who left against medical advice
8. Patients who expired
9. Patients discharged to home for hospice care
10. Patients discharged to a health care facility for hospice care
11. Patients with a documented Reason For Not Prescribing Statin Medication at Discharge

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
N=39,812

1. Patients < 18 years old = 0%
2. Patients who have a length of stay (LOS) greater than 120 days = 0%
3. Patients with Comfort Measures Only documented = 11.11%
4. Patients enrolled in clinical trials =0.39%
5. Patients admitted for Elective Carotid Intervention = 9.96%
6. Patients discharged to another hospital (acute care facility) = 1.13%
7. Patients who left against medical advice = 0.25%
8. Patients who expired = 3.58%
9. Patients discharged to home for hospice care = 0.60%
10. Patients discharged to a health care facility for hospice care = 1.52%
11. Patients with a documented Reason For Not Prescribing Statin Medication at Discharge = 2.58%

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Not Applicable

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:
Not Applicable

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:
Not Applicable

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:
Not Applicable

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:
The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
STK-6 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10th Percentile
3Q2011	152	0.92074	0.14322	1	1	0.96214	0.9	0.82353
2Q2011	155	0.90883	0.16942	1	1	0.95833	0.9125	0.77778
1Q2011	159	0.87312	0.18575	1	1	0.9375	0.84615	0.6
4Q2010	136	0.87766	0.17491	1	0.98626	0.93333	0.83667	0.69231
3Q2010	126	0.89971	0.17615	1	1	0.96	0.9	0.7
2Q2010	117	0.90979	0.15111	1	1	0.96667	0.88889	0.72727
1Q2010	97	0.9034	0.15012	1	1	0.95652	0.875	0.67273
4Q2009	48	0.88567	0.18656	1	1	0.94949	0.85165	0.66667

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Multiple data sources are not used for this measure.

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:
Not applicable

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:
Not applicable

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: The measure is not stratified for disparities.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:
Although some evidence exists that minorities receive statin therapy less frequently and aggressively than non-Hispanic whites, currently there are no plans to stratify the measure. The Joint Commission does not currently capture gender-specific data, or data elements for race or ethnicity because these data elements have not been shown to be reliably collectable due to the fact that no national standardized definitions exist for these data elements. Also, not all hospitals collect race and ethnicity. In the future, it may be feasible for The Joint Commission to explore how race and ethnicity and other relevant disparity data, might be collected reliably.

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met? (*Reliability and Validity must be rated moderate or high*) Yes☐ No☐ Provide rationale based on specific subcriteria:
If the Committee votes No, STOP

NEW (since 2012 endorsement): Data for Empirical Testing

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

2b2. VALIDITY TESTING
2b2.1. What level of validity testing was conducted? (*may be one or both levels*)
☐ **Critical data elements** (*data element validity must address ALL critical data elements*)
☐ **Performance measure score**
☐ **Empirical validity testing**
☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The data used to measure validity consists of one year data (third, and fourth quarter 2014 and first and second quarter for 2015) extracted from The Joint Commission warehouse. The hospitals selection was based to those hospitals that reported 12 months of data and had 30 or more denominator cases for the year.

Measure convergent and criterion validity was assessed using hospitals patient level data from The Joint commission warehouse. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous face validity testing. The patient population was comprised of all patients discharged with an ICD-9-CM Principal Diagnosis Code for stroke as defined in Appendix A, Table 8.1 or Table 8.2.

Conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications. The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015.

We conducted a secondary analysis to help interpret results; correlation with other stroke process measures.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Descriptive statistics for the STK measures:
N=1,318 hospitals
n= 2, 206,379 patients records

STK-6

Median: 100%
Percentile 10%: 92%
Percentile 25%: 96%
Percentile 75%: 100%
Percentile 90%: 100%

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations							
	STK_1	STK_2	STK_3	STK_4	STK_5	STK_6	STK_10
STK_1	1.00000 1318	0.55423 <.0001 1318	0.30311 <.0001 1302	0.37009 <.0001 1144	0.52785 <.0001 1318	0.53082 <.0001 1318	0.45291 <.0001 1318
STK_2	0.55423 <.0001 1318	1.00000 1318	0.37566 <.0001 1302	0.28169 <.0001 1144	0.53355 <.0001 1318	0.47326 <.0001 1318	0.29578 <.0001 1318
STK_3	0.30311 <.0001 1302	0.37566 <.0001 1302	1.00000 1302	0.25551 <.0001 1137	0.25665 <.0001 1302	0.28198 <.0001 1302	0.15819 <.0001 1302
STK_4	0.37009 <.0001 1144	0.28169 <.0001 1144	0.25551 <.0001 1137	1.00000 1144	0.22516 <.0001 1144	0.43717 <.0001 1144	0.41076 <.0001 1144
STK_5	0.52785 <.0001 1318	0.53355 <.0001 1318	0.25665 <.0001 1302	0.22516 <.0001 1144	1.00000 1318	0.35302 <.0001 1318	0.39079 <.0001 1318
STK_6	0.53082 <.0001 1318	0.47326 <.0001 1318	0.28198 <.0001 1302	0.43717 <.0001 1144	0.35302 <.0001 1318	1.00000 1318	0.45420 <.0001 1318
STK_10	0.45291 <.0001 1318	0.29578 <.0001 1318	0.15819 <.0001 1302	0.41076 <.0001 1144	0.39079 <.0001 1318	0.45420 <.0001 1318	1.00000 1318

The correlations of the stroke measures are all positively correlated and statistically significant indicating that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)
Overall, the positive inter-correlations indicates convergent validity of the measures. STK-06 is positively correlated with other stroke evidence-based processes of care.

2b3. EXCLUSIONS ANALYSIS .
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical*

analysis was used)

The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patient who have a Length of Stay greater than 120 days
- Patients with *Comfort Measures*
- Patient enrolled in a Clinical Trial
- Patients admitted for *Elective Carotid Intervention*
- Patients with a documented *Reason For Not Prescribing Statin Medication at Discharge*
- Patients discharged to another hospital
- Patients who left against medical advice
- Patients who expired
- Patients discharged to home for hospice care
- Patients discharged to a health care facility for hospice care

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are given below.

Exclusion: *Comfort Measures Only*
Overall Occurrence n = 163,225
Overall Occurrence Percentage: 10.4%
Minimum: 0.30%
10th Percentile: 5.26%
Median: 10%
90th Percentile: 15.8%
Maximum: 35.7%

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%
10th Percentile: 0.24%
Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for *Elective Carotid Intervention*
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.5%
90th Percentile: 25.3%
Maximum: 95.2%

Exclusion: Patients with a documented *Reason For Not Prescribing Statin Medication at Discharge*
Overall Occurrence n = 13,297
Overall Occurrence Percentage: 4.2%
Minimum: 0.24%
10th Percentile: 1.11%
Median: 3.31%

90th Percentile: 10.5%
Maximum: 29.7%

Exclusion: *Discharge Disposition* - Patients discharged to another hospital
Overall Occurrence n = 449,924
Overall Occurrence Percentage: 35.7%
Minimum: 0.787%
10th Percentile: 25%
Median: 35.4%
90th Percentile: 46%
Maximum: 76.2%

Exclusion: *Discharge Disposition* - Patients who left against medical advice
Overall Occurrence n = 8,396
Overall Occurrence Percentage: 0.67%
Minimum: 0.067%
10th Percentile: 0.34%
Median: 0.86%
90th Percentile: 2.4%
Maximum: 9.67%

Exclusion: *Discharge Disposition* - Patients who expired
Overall Occurrence n = 76,168
Overall Occurrence Percentage: 6.04%
Minimum: 0.398%
10th Percentile: 1.9%
Median: 5.08%
90th Percentile: 10.2%
Maximum: 20.1%

Exclusion: *Discharge Disposition* - Patients discharged to home for hospice care
Overall Occurrence n = 658,264
Overall Occurrence Percentage: 52.2%
Minimum: 6.25%
10th Percentile: 39%
Median: 51.9%
90th Percentile: 64%
Maximum 94.3%

Exclusion: *Discharge Disposition* - Patients discharged to a health care facility for hospice care
Overall Occurrence n = 37,804
Overall Occurrence Percentage: 3%
Minimum: 0.169%
10th Percentile: 0.87%
Median: 3.01%
90th Percentile: 6.4%
Maximum: 23%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The median frequency of exclusions range from low to moderate. The distribution of exclusions across

hospitals is generally narrow indicating that the occurrence is random and likely would not bias performance results, although the percentage of patients excluded may differ depending on whether the hospital was a stroke center.

For criterion validity, it is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a Length of Stay greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Patients with *Comfort Measures Only* documented
Rationale: It is inappropriate to treat such patients beyond those measures necessary for comfort.

Patients enrolled in a Clinical Trial
Rationale: Anticoagulation therapy as specified by this measure may compromise a clinical trial.

Patients admitted for *Elective Carotid Intervention*
Rationale: Patients who are not admitted for treatment of an acute stroke may not require statin therapy

Patients with a documented *Reason For Not Prescribing Statin Medication at Discharge*
Rationale: It is inappropriate to treat patients who have a documented reason or contraindication to statin therapy

Patients discharged to another hospital
Rationale: This measure is meant for patients discharged to home.

Patients who left against medical advice
Rationale: Hospitals do not have opportunity for provision of quality care for the non-compliant patient.

Patients who expired
Rationale: Patients who expire are not eligible to be in this measure

Patients discharged to home for hospice care
Rationale: Statin therapy may not be warranted for the hospice patient

Patients discharged to a healthcare facility for hospice care
Rationale: Statin therapy may not be warranted for the hospice patient.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other

If other: Data element allowable values are selected, either manually or electronically, from clinical and coded data available in medical record documentation. All medical record documentation is used in the abstraction process. Vendor data collection tools are used to import data elements needed for measure rate calculation.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract the electronic version of this measure, so continues to offer this chart-abstracted version which allows for data capture from unstructured data fields. All data elements needed to compute the STK-6 performance measure score have been retooled for capture from electronic sources. Annual updates are performed to match the eCQM specifications to the current version of the chart-abstracted specifications.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

At the present time, hospitals using this performance measure generally collect measure data via manual review of the paper medical record, the EHR or a combination of both. Collected data are submitted to The Joint Commission on a quarterly basis, by way of contracted performance measurement system vendors, as described previously. Specifications for this measure are freely available to anyone who wishes to use the measure. Feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As described above, as feedback from measure users has indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

There are no fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public

domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Quality Check® http://www.qualitycheck.org/consumer/searchQCR.aspx Hospital Compare https://www.medicare.gov/hospitalcompare/search.html Public Health/Disease Surveillance Paul Coverdell National Acute Stroke Registry http://www.cdc.gov/dhdsp/programs/stroke_registry.htm Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program http://www.jointcommission.org/ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Annual Report – Improving America's Hospitals http://www.jointcommission.org/annualreport.aspx Quality Improvement (Internal to the specific organization) Disease-Specific Care Certification for Comprehensive Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx Disease-Specific Care Certification for Primary Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

• Name of program and sponsor: Quality Check®; The Joint Commission

• Purpose: A public website that allows consumers to: search for accredited and certified organizations by city and state, by name or by zip code (up to 250 miles); find organizations by type of service provided within a geographic area; download free hospital performance measure results; and, print a list of Joint Commission certified disease-specific care programs and health care staffing firms.

• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-

accredited hospitals (2014)

• Name of program and sponsor: Hospital Compare; Centers for Medicare & Medicaid Services
• Purpose: A public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4000+ Medicare-certified hospitals (2015)

• Name of program and sponsor: Paul Coverdell National Acute Stroke Registry; Centers for Disease Control and Prevention
• Purpose: A national registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.
• Geographic area and number and percentage of accountable entities and patients included: 11 states; 403 hospitals (CDC, 2014)

• Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services
• Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 hospitals (2015)

• Name of program and sponsor: Annual Report-Improving America's Hospitals; The Joint Commission
• Purpose: The annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
•Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

• Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

• Name of program and sponsor Disease-Specific Care Certification for Primary Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
• Geographic area and number and percentage of accountable entities and patients included Nationwide; 1079 hospitals

• Name of program and sponsor Hospital Accreditation Program; The Joint Commission
• Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.
• Geographic area and number and percentage of accountable entities and patients included Nationwide; 3300 Joint Commission-accredited hospitals (2014)

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

• Progress (trends in performance results, number and percentage of people receiving high-quality healthcare):
Over the past five years, there has been steady but small increases of ~1-2% per year in the percentage of ischemic stroke patients with LDL > 100 mg/dL who are discharged on a statin medication based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 98.7%. A modest gap of ~5% still exists for the lowest decile of hospitals. This trend is consistent with increases in the rates for discharged on a statin medication published by GWTG-Stroke and PCNASR.

Racial and ethnic disparities persist for patients prescribed statin therapy at discharge, especially for black patients, as published based on data from the GWTG database (Schwamm LH, et al., 2010; Qian F, et al., 2013). Women are also less likely to receive a statin medication at discharge when compared to men (PCNASR, 2014).

• Geographic area and number and percentage of accountable entities and patients included:
Nationwide; 1296 hospitals; 140,296 patients (The Joint Commission, 2014)

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Although historic data for STK-6 demonstrates improvement over time for those ischemic stroke patients with LDL > 100 mg/dL, LDL not measured, or who were on a lipid-lowering medication prior to hospital arrival, there are no data currently available to evaluate performance for those ischemic stroke patients with LDL < 100 mg/dL who are prescribed a statin medication at discharge. Measure specifications have been revised to reflect current guideline recommendations. Data collection using the revised specifications was initiated with discharges on and after October 1, 2015. Somewhat lower measure rates are anticipated with the addition of those ischemic stroke patients with LDL < 100 mg/dL who are now included in the denominator population.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No unintended consequences due to the previous measure specifications or resulting from the recent changes to the measure specifications have been identified. One specific change to the measure specifications was modification of the Notes for Abstraction section for the data element definition Reason for Not Prescribing Statin Medication at Discharge. This change allows ischemic stroke patients with LDL < 70 mg/dL to be excluded from the measure population, and mitigates concerns about removal of the LDL data elements or the potential for inappropriate use of statins.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

0074 : Chronic Stable Coronary Artery Disease: Lipid Control
0118 : Anti-Lipid Treatment Discharge
0543 : Adherence to Statin Therapy for Individuals with Cardiovascular Disease
0545 : Adherence to Statins for Individuals with Diabetes Mellitus
0547 : Diabetes and Medication Possession Ratio for Statin Therapy
0639 : Statin Prescribed at Discharge
1519 : Statin Therapy at Discharge after Lower Extremity Bypass (LEB)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0543 : Adherence to Statin Therapy for Individuals with Cardiovascular Disease; Centers for Medicare and Medicaid Services
0545 : Adherence to Statins for Individuals with Diabetes Mellitus – measure not in NQF database
0547 : Diabetes and Medication Possession Ratio for Statin Therapy; CMS
0639 : Statin Prescribed at Discharge; CMS

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Three statin therapy measures were identified from the NQF database. All three measures address target diagnoses other than ischemic stroke or specific surgical procedures for patients 18 years or older: 0074 Coronary Artery Disease; 0118 isolated Coronary Artery Bypass Graft (CABG); and, 1519 Lower Extremity Bypass (LEB). Measure 1519 addresses inpatient organizational performance.. The other two measures, 0074 and 0118 are provider-level measures in the ambulatory care setting.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not Applicable

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0441
Measure Title: STK-10: Assessed for Rehabilitation
Measure Steward: The Joint Commission
Brief Description of Measure: This measure captures the proportion of ischemic or hemorrhagic stroke patients assessed for or who received rehabilitation services during the hospital stay.
This measure is a part of a set of eight nationally implemented measures that address stroke care (STK-1: Venous Thromboembolism (VTE) Prophylaxis, STK-2: Discharged on Antithrombotic Therapy, STK-3: Anticoagulation Therapy for Atrial Fibrillation/Flutter, STK-4: Thrombolytic Therapy, STK-5: Antithrombotic Therapy By End of Hospital Day 2, STK-6 Discharged on Statin Medication, and STK-8: Stroke Education) that are used in The Joint Commission's hospital accreditation and Disease-Specific Care certification programs.
Developer Rationale: Stroke is a leading cause of serious, long-term disability, associated with significant costs. The primary goal of rehabilitation is to prevent complications, minimize impairments, and maximize function. Evidence suggests that better clinical outcomes are achieved when post-acute stroke patients receive coordinated, multidisciplinary evaluation and intervention through the provision of rehabilitation services. Therefore, the need for rehabilitation services after an ischemic or hemorrhagic stroke should be assessed prior to discharge from the acute care setting.
Healthcare organizations that track rehabilitation assessments for internal quality improvement purposes have seen an improvement in the measure rate over time. This measure has been included in the Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

Numerator Statement: Ischemic or hemorrhagic stroke patients assessed for or who received rehabilitation services.
Denominator Statement: Ischemic or hemorrhagic stroke patients.
Denominator Exclusions: • Less than 18 years of age
• Length of Stay > 120 days
• Comfort measures only documented
• Enrolled in clinical trials related to stroke
• Admitted for elective carotid intervention
• Discharged to another hospital
• Left against medical advice
• Expired
• Discharged to home for hospice care
• Discharged to a health care facility for hospice care

Measure Type: Process
Data Source: Electronic Clinical Data, Paper Medical Records
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Jul 31, 2008 **Most Recent Endorsement Date:** Nov 01, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Summary of prior review in 2012

- Evidence suggests that better clinical outcomes are achieved when post-acute stroke patients receive coordinated, multidisciplinary evaluation and intervention through the provision of rehabilitation services. Therefore, the need for rehabilitation services after an ischemic or hemorrhagic stroke should be assessed prior to discharge from the acute care setting.
- AHA/ASA-Endorsed Management of Adult Stroke Rehabilitation Care: A Clinical Practice Guideline (2005).
    - Level I Evidence, Recommendation A - Organized and coordinated post-acute inpatient rehabilitation care improves outcome (page e-120).
        - Strongly recommend that once the patient is medically stable, the primary physician consult rehabilitation services (i.e., physical therapy, occupational therapy, speech and language pathology, kinesiotherapy, and physical medicine), as indicated, to assess the patient's rehabilitation needs and to recommend the most appropriate setting to meet those needs (page e-119).
- The 2012 Committee expressed no concerns regarding the evidence underlying this measure.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☐ **The developer provided updated evidence for this measure:**

Exception to evidence

Not applicable

Guidance from the Evidence Algorithm

Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: Moderate; Consistency: Moderate/High(Box 5) → Box 5b Moderate

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and voting on Evidence?*

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provides the following national tend data:

	CY 2010	CY2011	CY 2012	CY 2013	CY 2014
# hospitals	138	157	158	262	1299
# cases (den)	23,875	28,415	28,735	44,242	210,075
National aggregate rate	0.96	0.97	0.98	0.98	0.98
Mean hospital rate	0.95	0.96	0.96	0.97	0.97
50th percentile	0.97	0.98	0.98	0.99	0.99
10th and 90th percentiles	0.87 (10th) 1.00 (90th)	0.89 (10th) 1.00 (90th)	0.92 (10th) 1.00 (90th)	0.95 (10th) 1.00 (90th)	0.94 (10th) 1.00 (90th)

- Participation in this measure has grown significantly in the past five years. National hospital performance seems to be very high with little to no room for improvement.
- In 2012, the Committee noted the overall high rate of performance for this measure (approximately 96% among reporting hospitals).

Disparities

- The developer does not provide disparities data from use of this measure.
- Several references are cited "According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age."

Questions for the Committee:

- *Data for 2014 includes 1,299 hospitals. What proportion of hospitals caring for stroke patients is captured in this data?*
- *Data over five years show significant improvement. How much further improvement in performance is likely using this measure?*
- *How can this measure be used to better understand disparities in care and outcome for certain groups?*

Preliminary rating for opportunity for improvement: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **the evidence is the same as before, agree with assessment; yet despite this, the some denominator criteria have been

changed...do we need to reconcile?
**No new studies/information
**The evidence for the measure has not changed since the last NQF endorsement review thus there is no need for repeat discussion and voting on Evidence.
**No new evidence since 2012. No vote on evidence. There is new evidence, but it doesn't appear to change any conclusions. There is a strong need for evidence with regard to disparities and more specific details about relationship of assessment/rehabilitation and outcomes.

1b. Performance Gap

Comments: **There remains little room for improvement at this point; the one caveat might be that presenting data broken down by race and SES might suggest there are still pockets where work needs to be done; not clear if developer has access to that data (via partnership with GWTG?). The literature presented suggest there are indeed still gaps by gender and race...rating might not be low if you can identify underserved groups where a meaningful gap still exists.
**high participation in this measure 98%
**Data on disparities is not provided, however there are many studies referenced that illustrate disparities with rehabilitation services based on SES and race/ethnicity. If outcomes from this measure could be micro-segmented to understand current gaps it would help address health inequities.
**Participation has grown considerably. Room for improvement is unclear. Without stratification of the data by various subgroups it is impossible to tell if there are notable performance gaps in various groups.

1c. High Priority (previously referred to as High Impact)

Comments: **NA
**N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

- One data element defines the numerator and eight data elements define the denominator.
- Developer reports that "Since the last endorsement date, there have been no significant changes to STK-10 Assessed for Rehabilitation. Minor clarifications regarding the composition of the rehabilitation team have been made, and advanced practice nurse (APN) and physician assistant (PA) added to physician documentation as acceptable data sources for a rehabilitation assessment. Clarification was needed to prevent inclusion of other types of clinical assessments, e.g. activities of daily living, fall risk, etc., in the numerator population."
- Diagnosis and procedure codes have been converted from ICD-9 to ICD-10-CM diagnosis and ICD-10-PCS procedure codes. See attached spreadsheet Appendix A Table 8.1 (ischemic stroke) and Table 8.2 (hemorrhagic stroke)
- A conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications:
    - "The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were

published in early 2015, well in advance of the mandated date of October 1, 2015 developer advisory panel determined that the intent of the measure was changed by the conversion to ICD-10 and a crosswalk was posted."

- Sampling monthly or quarterly is allowed; instructions for sampling are provided.
- A web-based data collection tool has been developed but is not described. Hospitals are not required to use this tool.
- The measures is specified at the hospital level of analysis.

Questions for the Committee :

- o *Are all the data elements clearly defined? Are all appropriate codes included?*
- o *Is the logic or calculation algorithm clear?*
- o *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

Reliability testing from prior review in 2012:

- Inter rater-reliability of the data elements for one year (4Q2010-3Q2011) in 77 hospitals and 739 patient records showed an overall agreement rate of 98.3%.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing see above

Results of reliability testing see above

Guidance from the Reliability Algorithm

Precise specifications (Box 1) → empirical testing as specified (Box 2) → reliability testing with patient-level data elements (Box 8) → assessed all critical data elements (Box 9) → high/moderate certainty the data elements are reliable (Box 10) → Moderate (NOTE: As testing was only done at the data element level and not at the measure score level, the highest rating possible is moderate.)

Questions for the Committee:

- o *The developer has not provided any new reliability testing. Does the Committee agree there is no need for repeat discussion and voting on Reliability?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

Validity testing from the prior review in 2012:

- Face validity of the data was assessed via survey and focus groups of hospitals participating in the pilot test.
- The developer reports that "All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant."
- The information provided does not indicate that face validity of the measure score as a representation of quality was systematically assessed.
- Ongoing feedback from measure users is systematically reviewed to identify trends and/or measure specifications that require clarification or revision. Predominantly, questions involving the data element Assessed for Rehabilitation Services, specifically questions regarding the composition of the rehabilitation team and the qualifications needed to complete the rehabilitation assessment despite detailed inclusion guidelines.
-

Describe any updates to validity testing– New empirical validity testing data is presented

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- Pearson Correlation Coefficients were calculated for seven stroke measures for hospitals that reported 12 months of data and had 30 or more denominator cases for the year (2014-2015).Presumably high performing hospitals would do well on all stroke performance measures.

Validity testing results:

- Analysis of 1,318 hospitals and 2,206,379 patients records generated a table of Pearson Correlation Coefficient results that shows a statistically significant (P<.0001), positive correlation of this measure with six other stroke performance measures supporting the hypothesis that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- Patients are excluded from the measure for the following reasons:
  - Patients less than 18 years of age
  - Patient who have a Length of Stay greater than 120 days

- Patients with Comfort Measures
- Patient enrolled in a Clinical Trial
- Patients admitted for Elective Carotid Intervention
- Patients discharged to another hospital
- Patients who left against medical advice
- Patients who expired
- Patients discharged to home for hospice care
- Patients discharged to a health care facility for hospice care

- New data on the frequency of exclusions is presented for 2, 206,379 admissions in 1,318 hospitals:
  - Comfort Measures Only: 10% (range 0.30-35.7%)
  - Clinical Trial: 0.52% (range 0.08- 10%)
  - Patients admitted for Elective Carotid Intervention: 11.5% (range 0.16-95.2%)
  - Patients discharged to another hospital: 35.4% (range 0.78 -76.2%)
  - Patients who left against medical advice: 0.86% (range 0.06-9.67%)
  - Patients who expired: 5.08% (range 0.39-20.1%)
  - Patients discharged to home for hospice care: 51.9% (range 6.25-94.3%)
  - Patients discharged to a health care facility for hospice care: 3% (range 0.16-23%)
  -

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Performance data is presented above under opportunity for improvement. Complete details of the data is presented in 1b.
- The developer explains their approach: "There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations."

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods: Not applicable

2b7. Missing Data

- The developer describes the handling of missing data in the measure specifications (S.22) but does not provide information on the frequency of missing data or potential impact on measure results.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → potential threats to validity mostly assessed (Box2) →empirical validity testing (Box 3) → measure score testing (Box 6) → sound conceptualization and hypotheses (Box 7) →moderate confidence that score are a valid indicator of quality(Box 8a) → High

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **changed denom specs to agree with 2013 ACC, but then ignore the ACC <75 recommendation? Also, added LDL <70 as reason not to give, new but no new evidence to support

**Specifications clear

**The test sample is adequate to generalize for widespread implementation. The results demonstrate sufficient validity so that conclusions about quality can be made. The score from this measure as specified is an indicator of quality.

**Again, validity appears to be quite strong within the data presented. Validity with regard to various population groups and various assessments and rehabilitation methods in relation to outcomes would be quite helpful.

2a2. Reliability Testing

Comments: **o Is the test sample adequate to generalize for widespread implementation? Very large sample, so probably

o Do the results demonstrate sufficient validity so that conclusions about quality can be made? I think so

o Do you agree that the score from this measure as specified is an indicator of quality? yes

**New testing of measure is sufficient

**The test sample is adequate to generalize for widespread implementation. The results demonstrate sufficient validity so that conclusions about quality can be made. The score from this measure as specified is an indicator of quality.

**Validity is also quite good within the existing framework - stretching beyond that framework, as above would be desirable.

**New criteria testing not seen since so new?

**No new testing needed

2b2. Validity Testing

Comments: **Home for hospice 52% as largest exclusion, doesn't sound right.

As discussed above, not sure the differences at hospital level have much meaning

**large# pts excluded in new data, eg, 51.9% of pts discharged to home for hospice

**Information is not provided on missing data frequency nor impact on results. This may constitute a threat to the validity of this measure.

**High levels of exclusions are a concern. It seems likely some of those groups that are excluded might be worthy of inclusion.

**Is the test sample adequate to generalize for widespread implementation? Very large sample, so probably

Do the results demonstrate sufficient validity so that conclusions about quality can be made? I think so

Do you agree that the score from this measure as specified is an indicator of quality? yes

**New testing of measure is sufficient

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **new criteria testing not seen since so new?

**No new testing needed

**There is no need for repeat discussion and voting on Reliability.

**Reliability testing within the framework is quite strong - voting may not be necessary.

**"Home for hospice 52% as largest exclusion, doesn't sound right.

As discussed above, not sure the differences at hospital level have much meaning"

**Large # pts excluded in new data, eg, 51.9% of pts discharged to home for hospice

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is medical record abstraction: not all hospitals are able to generate the data electronically so a paper-based option is available
- Some data elements are in defined fields in electronic sources.
- Data collection burden is significant.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **quite feasible

**No concerns

**The required data elements are routinely generated and used during care delivery. The required data elements are available in electronic form. The data collection strategy is ready to be put into operational use.

**Data are available in the records (EHR in 2837). Should be relatively easy to use. There may be cases where the type of assessment and the type of rehabilitation are variable making interpretation less clear.

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details

- The Joint Commission Quality Check®: Public website that allows consumers to search for accredited and certified organizations.
- CMS Hospital Compare: Public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.
- CDC Paul Coverdell National Acute Stroke Registry: National registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke

care within stroke systems of care.

- CMS Hospital Inpatient Quality Reporting Program (IQR): The Hospital IQR program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.
- The Joint Commission Annual Report-Improving America's Hospitals: Annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.
- The Joint Commission Disease-Specific Care Certification for Comprehensive Stroke Centers: Certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
- The Joint Commission Disease-Specific Care Certification for Primary Stroke Centers: Certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.
- The Joint Commission Hospital Accreditation Program: Accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

Improvement results

- The developer provided the following progress information on 1,256 hospitals nationwide and 180,048 patients:
    - The rate of patients assessed for rehabilitation services during the inpatient hospitalization has steadily increased ~1% per year over the past five years based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 98.7%. A modest gap of approximately 5% still exists for the 10th percentile of hospitals. This trend is consistent with increases in the rate in the assessment for rehabilitation services published by GWTG-Stroke and PCNASR.
    - Significant and important differences between races in rates of length of stay and less likely to be discharged home. (Qian, 3013; CDC 2014).

Unexpected findings (positive or negative) during implementation:

- Developer did not identify unexpected findings during implementation.

Potential harms:

- The developer did not identify any unintended consequences related to this measure.

Feedback :

- The developer states that feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As previously noted, measure users have indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **adequate

**High usability

**Although small, there are still opportunities to improve in this measure. No unintended consequences were identified. Identifying disparities in care by SES or race/ethnicity as above would contribute to improved quality of care.

**Data will be public and can easily be used for accountability. They will also provide a foundation for further research into some of the subgroups which may have varied results.

Criterion 5: Related and Competing Measures

Related or competing measures

NQF # 0244: Stroke and Stroke Rehabilitation: Rehabilitation Services Ordered- *no longer NQF-endorsed*

Harmonization:

Not applicable

Pre-meeting public and member comments

No comments were submitted for this measure during the pre-meeting comment period.

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0441 NQF Project: Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.
Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The focus of the measure is to prevent complications, minimize impairments, and maximize function post stroke by assessing the patient's need for rehabilitation services and/or initiating rehabilitation.
Assessed for rehabilitation need >> rehabilitation services received >> improved neurological outcomes >> decreased morbidity and mortality.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline
Selected individual studies (rather than entire body of evidence)
Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
This measure is consistent with the adult stroke rehabilitation care guidelines developed for the Veterans Health Administration and Department of Defense Medical Systems and endorsed by the American Heart Association/American Stroke Association (2005). Guideline recommendations incorporated information from existing evidence-based guidelines from the Agency for Health Care Policy and Research (AHCPR) Post-Stroke Rehabilitation (1995), the Scottish Intercollegiate Guidelines Network (SIGN) Management of Patients with Stroke (1997), and the Royal College of Physicians (RCP) National Clinical Guidelines for Stroke (2000).

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: According to the 2011 update from the Evidence-Based Review of Stroke Rehabilitation (EBRSR), 2000 studies including 1, 078 randomized control trials (RCTs) have been identified in the stroke rehabilitation literature. The 2001 EBSRS identified five meta-analyses of the effectiveness of stroke rehabilitation:

- Langhorne, et al. (1993) – 10 RCTs conducted between 1962 and 1993. Results revealed that management of stroke patients on a stroke unit was associated with lower mortality rates than general medical wards, and a 28% reduction in the risk of death occurring in the first 17 weeks post-stroke.
- Ottenbacher and Jannell (1993) – 36 RCTs (n=3,717 patients). Patients who participated in an individualized stroke rehabilitation programs performed better than 65% of those patients in the comparison group. Greater functional improvements were observed in younger patients and those with relatively short stroke onset to rehabilitation admission intervals.
- Stroke Unit Trialists' Collaboration (2007) - 31 RCTs. Stroke unit care was associated with a significant reduction in death (OR 0.86; 95% CI 0.76-0.98; P=0.02) at a median of one-year follow-up. Stroke unit care was also associated with a significant reduction in the combined outcomes or both death or institutional care (OR 0.82; 95% CI 0.73-0.92; P=0.0006) and death or dependency (OR 0.82; CI 95% 0.73-0.92, P=0.001).
- The Canadian Coordinating Office of Health Technology Assessment (CCOHTA) (2003) – 6 RCTs (n=1,709 patients) from 1995 to July 2002. Stroke unit care was associated with a reduction in the odds of death (OR 0.60; CI 95% 0.42-0.86) an outcome recorded in all studies.
- Seenan, et al. (2007) – 18 non-randomized trials, which more closely approximate usual clinical practice. Findings reported

were similar to prior meta-analyses of RCTs. The odds of death (OR 0.79; 95% CI 0.72-0.86) and poor outcome (OR 0.87; CI 9%%0.80-0.95) were reduced for patients receiving stroke unit care compared to general medical management.

In addition, a total of 37 individual studies (14 non-randomized and 23 RCTs) of the efficacy of stroke rehabilitation were included in the review. Only the results of RCTs and quasi RCTs were used to formulate conclusions. Studies were categorized according to the type of care provided ranging from acute stroke management to settings that offered various levels of rehabilitation services (i.e., combined acute and rehabilitation stroke units, subacute rehabilitation units, and mobile stroke teams). Overall, the body of evidence strongly supports that units providing stroke rehabilitation are associated with improved functional outcomes for patients.

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The quality of evidence supporting stroke rehabilitation is moderate. The impact of stroke rehabilitation on outcomes has been difficult to quantify due to problems with study design and methodology (e.g., lack of randomization, inappropriate control group selection, failure to blind assessors, difficulty in controlling for all possible confounders) detected by systematic review. Furthermore, issues inherent to stroke rehabilitation, such as controlling for spontaneous neurological recovery, daily fluctuation in individual function, and difficulties measuring functional outcomes have challenged study designs. Pre-selection of patients and observer measurement bias are additional concerns when studying the impact of stroke rehabilitation on outcomes (Foley, et al, 2011).

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* The body of evidence consistently supports that rehabilitation services minimize functional impairment and improve clinical outcomes for patients following stroke. Furthermore, most stroke patients have some residual deficits following the event, and can benefit from rehabilitation services tailored to improved cognitive, speech, or motor function. Based on these findings, a rehabilitation assessment should be completed for all stroke patients prior to hospital discharge.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
There is strong evidence on the net benefits of stroke rehabilitation. The body of evidence clearly supports that stroke rehabilitation is associated with a reduction in the odds of death or dependency and the need for institutionalization. There is strong evidence that stroke rehabilitation improves functional outcomes. There is also strong evidence that stroke rehabilitation reduces mortality for a subset of more severe stroke patients.

There is no dispute that stroke represents a significant economic burden in developed countries; however, there is some uncertainty respecting the cost-effectiveness of stroke rehabilitation. Most of the studies have been conducted outside of the United States in the United Kingdom, Australia, or the Netherlands. Stroke recovery and residual disability are highly variable. Studies have ignored the contributions of family members and/or caregivers towards rehabilitative progress, and the discrete components of caregiver and associated costs are difficult to isolate.

Although findings differ, there is some evidence of the cost-effectiveness of stroke rehabilitation services. Kalra et al. (2005) found stroke unit care to be more effective than home care, and also to be of equal cost (using per patient day alive), suggesting that stroke unit care is more cost-effective than home care. Using prospectively collected data from the Stroke Care Outcome: Providing Effective Services (SCOPES) trial over six months, Moodie et al. (2007 - Australia) compared the effectiveness of stroke units, conventional care, and a mobile service. While better outcomes were noted for patients cared for in stroke units, the incremental costs were higher compared to conventional care. Saka et al. (2009 – UK) projected the cost-effectiveness of three types of care over a 10-year period: stroke units with early supported discharge (ESD), stroke units without ESD, and general medical-surgical care. Although the costs of care were greater for both stroke units when compared to general medical-surgical wards, the cost per quality-adjusted life years (QALY) was lowest for stroke units with ESD (Foley, et al., 2011).

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: The Evidence-Based Review of Stroke Rehabilitation (EBRSR) funded by the Canadian Stroke Network (CSN).

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: Canadian Stroke Network; 2010 Dec 8. p. 129-150.
Rating Scheme for the Strength of the Evidence
Summary of Definitions for Levels of Evidence*
Grade A – Criteria – Strong recommendation. Evidence from randomized controlled trials or meta-analysis of randomized controlled trials. Desirable effects clearly outweigh undesirable effects, or vice versa.
Grade B – Criteria – Single randomized controlled trial or well-designed cohort or case-control analytic study; or multiple time series or dramatic results of uncontrolled experiment. Desirable effects closely balanced with undesirable effects.
Grade C – Criteria – At least one well-designed, nonexperimental descriptive study (e.g., comparative studies, correlation studies, case studies) or expert committee reports, opinions and/or experience of respected authorities, including consensus from development and/or reviewer groups.
*Based on Guyatt GH, Cook DJ, Jaeschke R, et al. Grades of recommendation for antithrombotic agents: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th edition). [published erratum in Chest 2008;34:47]. Chest 2008;133(6 Suppl):123S-131S.

1c.13 Grade Assigned to the Body of Evidence: Grade A Strong Evidence

1c.14 Summary of Controversy/Contradictory Evidence: The benefits of stroke rehabilitation are obvious. No controversies over the benefits of stroke rehabilitation were noted in the literature.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

- Foley N, Teasell R, Bhogal S, Speechley M. The efficacy of stroke rehabilitation. The Evidence-Based Review of Stroke Rehabilitation. August 2011: 1-50.
- Kalra L, Evans A, Perez I, Knapp M, Swift D, Donaldson N, Swift CG. Alternative strategies in stroke care. Health Technology Assessment. 2005;9:1-94.
- Keith RA. Rehabilitation after stroke: cost-effectiveness analyses. Journal of the Royal Society of Medicine. 1996;89:631-633.
- Langhorne P, Williams BO, Gilchrist W, Howie K. Do stroke units save lives? Lancet. 1993;342:395-398.
- Moodie M, Cadilhac D, Pearce D. Economic evaluation of Australian stroke services: a prospective, muticenter study comparing dedicated stroke units with other care modalities. Stroke. 2006:37:2790-2795.
- Noorani HZ, Brady B, McGahan L, Teasell R, Skidmore B, Doherty T. Stroke rehabilitation services: systematic reviews of the clinical and economic evidence. Canadian Coordinating Office for Health Technology Assessment. March 2003; Technology Report No. 35.
- Ottenbacher KJ, Jannell S. The results of clinical trials in stroke rehabilitation research. Arch Neurol. 1993;50:37-44.
- Saka O, Serra V, Samyshkin Y, McGuire A, Wolfe CC. Cost-effectiveness of stroke unit care followed by early supported discharge, Stroke. 2009;40:24-49.
- Zorowitz RD, et al. the Post-Stroke Rehabilitation Outcomes Project (PSROP), Top Stroke Rehabil. 2005 Fall;12(4).

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)*:
AHA/ASA-Endorsed Management of Adult Stroke Rehabilitation Care: A Clinical Practice Guideline (2005).
Level I Evidence, Recommendation A
Organized and coordinated post-acute inpatient rehabilitation care improves outcome (page e-120).
1. Strongly recommend that once the patient is medically stable, the primary physician consult rehabilitation services (i.e., physical therapy, occupational therapy, speech and language pathology, kinesiotherapy, and physical medicine), as indicated, to assess the patient's rehabilitation needs and to recommend the most appropriate setting to meet those needs (page e-119).

1c.17 Clinical Practice Guideline Citation: Duncan PW, Zorowitz R, Bates B, Choi JY, Glasberg JJ, Graham GD, Katz RC, Lamberty K, Refer D. Management of Adult Stroke Rehabilitation Care: A Clinical Practice Guideline. Stroke. 2005;36:e100-143.

1c.18 National Guideline Clearinghouse or other URL: http://stroke.ahajournals.org/content/36/9/e100.full.pdf+html?sid=b27cb2e0-836f-4fc0-9ee0-aae14d24f3e4

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation

and any disclosures regarding bias: The Stroke Council of the American Heart Association has chosen to endorse the VA DoD guideline for stroke rehabilitation (Management of Stroke Rehabilitation. Washington, DC: VA/DoD Clinical Practice Guideline Workgroup, Veterans Health Administration, Department of Veterans Affairs and Health Affairs, Department of Defense, February 2003). The Department of Defense and Veterans Health Administration panel of experts evaluated the medical evidence according to criteria proposed by the US Preventive Services Task Force. The American Heart Association/American Stroke Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

1c.21 System Used for Grading the Strength of Guideline Recommendation: USPSTF

1c.22 If other, identify and describe the grading scale with definitions:

1c.23 Grade Assigned to the Recommendation: A: A strong recommendation that the intervention is always indicated and acceptable.

1c.24 Rationale for Using this Guideline Over Others: The American Heart Association (AHA) is the leading organization in the United States that fosters appropriate cardiac care in an effort to reduce disability and deaths caused by cardiovascular disease and stroke. The American Stroke Association (ASA) is an AHA affiliated organization which focuses on care, research, and prevention of stroke. AHA/ASA Scientific Statements and clinical guideline recommendations provide physicians, emergency healthcare providers, and other stroke care professionals with current evidence on components of the evaluation and treatment of stroke patients. AHA/ASA statements and recommendations are released and updated on a continuing basis.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?
1c.25 Quantity: High 1c.26 Quality: Moderate1c.27 Consistency: Moderate

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0441_Evidence_MSF5.0_Data.doc

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is a leading cause of serious, long-term disability, associated with significant costs. The primary goal of rehabilitation is to prevent complications, minimize impairments, and maximize function. Evidence suggests that better clinical outcomes are achieved when post-acute stroke patients receive coordinated, multidisciplinary evaluation and intervention through the provision of rehabilitation services. Therefore, the need for rehabilitation services after an ischemic or hemorrhagic stroke should be assessed prior to discharge from the acute care setting.

Healthcare organizations that track rehabilitation assessments for internal quality improvement purposes have seen an improvement in the measure rate over time. This measure has been included in the Hospital Inpatient Quality Reporting Program for three years (i.e., FY 2015, FY 2016, FY 2017) to promote improvements in quality at the national level.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
In October, 2009, The Joint Commission added the stroke (STK) measure set as a new core measure option to meet performance measure requirements for Joint Commission hospital accreditation purposes. Below is the specified level of analysis for STK-10 Assessed for Rehabilitation beginning with discharges 4Q 2009 through December 31, 2014.

4Q 2009: 2555 denominator cases; 2473 numerator cases; 49 hospitals; 0.96791 national aggregate rate; 0.95212 mean of hospital rates; 0.07659 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.97744 50th percentile rate/median rate; 0.95 25th percentile rate/lower quartile; and, 0.85294 10th percentile rate.

CY 2010: 23,875 denominator cases; 23,144 numerator cases; 138 hospitals; 0.96938 national aggregate rate; 0.95282 mean of hospital rates; 0.06331 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.97731 50th percentile rate/median rate; 0.93333 25th percentile rate/lower quartile; and, 0.875 10th percentile rate.

CY 2011: 28,415 denominator cases; 27,731 numerator cases; 157 hospitals; 0.97593 national aggregate rate; 0.96355 mean of hospital rates; 0.06802 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.98671 50th percentile rate/median rate; 0.96203 25th percentile rate/lower quartile; and, 0.89939 10th percentile rate.

CY 2012: 28,735 denominator cases; 28,272 numerator cases; 158 hospitals; 0.98389 national aggregate rate; 0.96919 mean of hospital rates; 0.06586 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.98917 50th percentile rate/median rate; 0.96875 25th percentile rate/lower quartile; and, 0.92857 10th percentile rate.

CY 2013: 44,242 denominator cases; 43,527 numerator cases; 262 hospitals; 0.98384 national aggregate rate; 0.97722 mean of hospital rates; 0.05526 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99131 50th percentile rate/median rate; 0.97605 25th percentile rate/lower quartile; and, 0.95652 10th percentile rate.

CY 2014: 210,075 denominator cases; 207,311 numerator cases; 1299 hospitals; 0.98684 national aggregate rate; 0.9772 mean of hospital rates; 0.06807 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.99433 50th percentile rate/median rate; 0.98 25th percentile rate/lower quartile; and, 0.94737 10th percentile rate.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
Not applicable

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Not applicable

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
According to a 2011 report from the American Heart Association/American Stroke Association, racial disparities in stroke care exist and are more predominant among people < 65 years of age. Evidence of disparities in stroke care between minority groups and whites include: lack of knowledge about the risk factors for stroke; lack of awareness about stroke signs and symptoms and the need for urgent treatment; and, access to care respecting prevention services, acute stroke treatment, and rehabilitation. Differences in care are also related to the socioeconomic status of minorities, insurance coverage, cultural beliefs and attitudes, language barriers, immigration status, mistrust of the healthcare system, and the number of providers representing minority groups. These are all factors contributing to the quality of stroke care (Cruz-Flores, et al. 2011).

Each year in the United States, ˜ 55,000 more women than men have a stroke. Statistics reveal that women have a higher "life-time risk of stroke" than men. (Mozaffarian D, et al., 2015). In the Framingham Heart Study, lifetime risk of stroke among those 55 to75 years of age was 1 in 5 for women (20% to 21%) and ˜1 in 6 for men (14% to 17%).

The burden of stroke is higher in Blacks or African Americans and Hispanics than whites. Racial and ethnic minorities have excess deaths from stroke and also experience greater years of potential life lost than non-Hispanic whites. The risk ratio for stroke mortality in all racial and ethnic minorities is higher in the 35-to-64-year-old age group, however, this risk decreases as people age. After age 64 non-Hispanic whites have an equal risk for stroke when compared to Hispanics and American Indian-Alaskan Natives. This equalization of rate of stroke presents again after age 85 in blacks or African Americans (Cruz-Flores, et al., 2011).

In the national REGARDS cohort, 27,744 black and white men and women, aged > 45 years, followed over 4.4 years, and stroke-free at baseline, reported an overall age-adjusted and sex-adjusted black/white incidence rate ratio of 1.51. At ages 45 to 54 years, the rate ratio increased to 4.02 compared to 0.86 for > 85 years. A higher incidence of stroke is reported for blacks at younger ages.

The REGARDS investigators found that approximately half of racial disparity in stroke risk is attributable to traditional risk factors (primarily systolic blood pressure) and socioeconomic factors (Howard, et al., 2011). Brown and colleagues (2011) found a higher incidence of ischemic stroke in disadvantaged white neighborhoods, but found no significant associations between neighborhood socioeconomic status and ischemic stroke among blacks. A recent large population-based Canadian study examined gender-adjusted, age-adjusted prevalence of cardiovascular risk factors, heart disease and stroke in four ethnic groups: white (n=154,653); South Asian (N=3364); Chinese (n=3038) and, blacks (n=2742). Stroke incidence was highest in the South Asian group (1.7%) and lowest in the Chinese population (0.6%). The increased risk in the South Asian population was attributed to high susceptibility to insulin resistance and metabolic syndrome, and a tendency to develop diabetes mellitus at younger ages in both men and women as compared to other ethnic groups (Chiu, et al., 2010).

The BASIC (Brain Attack Surveillance in Corpus Christi) project (NINDS) demonstrated an increased incidence of stroke among Mexican Americans compared with non-Hispanic whites in a community in southeast Texas. The crude 3-year cumulative incidence (2000-2003) was 16.8 per 1000 in Mexican Americans and 13.6 per 1000 in non-Hispanic whites. Specifically, Mexican Americans had a higher cumulative incidence for ischemic stroke at younger ages (45-59 years of age: RR 2.04, 95% CI 1.55-2.69; 60-74 years if age: RR 1.58, 95% CI 1.31-1.91) but not at older ages (> 75 years of age: RR 1.12, 95% CI 0.94-1.32). Mexican Americans also had a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

Temporal trend data from the BASIC Project for the time period 2000 through 2010 demonstrated that ischemic stroke rates declined significantly in people aged =60 years but remained largely unchanged over time in those aged 45 to 59 years. Rates of decline did not differ significantly for non-Hispanic whites and Mexican Americans in any age group. Therefore, ethnic disparities in stroke rates in the 45- to 59-year-old and 60- to 74-year-old age groups persist (Morgenstern, et al., 2013).

Data from the most recent Greater Cincinnati Northern Kentucky Stroke Study (GCNKSS) show that compared with the 1990s, when incidence rates of stroke were stable, stroke incidence in 2005 was decreased for whites. A similar decline was not seen in blacks. These changes for whites were driven by a decline in ischemic strokes. There were no changes in incidence of ischemic stroke for blacks or of hemorrhagic strokes in blacks or whites (Kleindorfer, et al., 2010). In an analysis of temporal trends in ischemic stroke incidence stratified by age, the GCNKSS found an increased incidence of ischemic stroke over time for both blacks and whites aged 20 to 54 years, especially in 2005 compared with earlier time periods. There were declining incidence rates in the oldest age groups for both race groups (Kissela, et al., 2012).

Few studies have examined racial-ethnic disparities in stroke rehabilitation. Findings from the studies available are somewhat contradictory. With regard to access, a study from the Veterans Health Administration found no differences in referral to or receipt of inpatient rehabilitation (Horner RD, et al., 2003); however, another study found that urban-dwelling blacks or African Americans were more likely to be discharged from the hospital to inpatient rehabilitation facilities than whites (Gregory PC, et al, 2006). The VA study from Horner also reported that blacks or African Americans had a slightly longer time to initiation of rehabilitation that whites (4.4 days versus 3.8 days; $P<0.05$).

Other studies have evaluated the effect of the Functional Independence Measure (FIM) score on the admission of minorities to inpatient rehabilitation facilities. No consistency of findings has been noted. A retrospective cohort study from Bhandari, et al. (2005) found no racial differences in admission FIM scores, but several other studies reported significant differences. One large study involving urban participants hospitalized at a county hospital reported that Hispanics had lower FIM scores than blacks or African Americans (Chiou-Tan FY, et al, 2006). Another study reported similar findings with whites having the highest FIM scores (Ottenbacher KJ, et al., 2008).

Rehabilitation length of stay has been studied with similarly inconsistent findings. One study of VA and non-VA inpatient rehabilitation facilities reported significantly longer lengths of stay for blacks or African Americans than for whites (Stineman MG, 2001). Another study from CDC (2007) using data from the BRFSS found that blacks or African Americans were more likely to be referred to outpatient rehabilitation services than whites (adjusted OR 1.49; 95% CI 1.1 to 2.0).

Results on the delivery of outpatient occupational and physical therapy are also conflicting with some reporting that whites are more likely to receive therapy than non-white minorities (Mayer-Oakes SA, 1992). Another study utilizing data from the Health and Retirement Study found no racial differences among patients receiving occupational or physical therapy (Cook C, 2005). Studies from Bhandari (2005), as well as Horner (1997), reported no racial differences in the intensity of occupational or physical therapy delivered to minorities when compared to whites.

A more recently published study (Qian F, et al, 2013) from the American Heart Association/American Stroke Association Get With The Guidelines (GWTG)-Stroke program from April 2003 through December 2008, (n=200,900), found racial disparities related to length of hospitalization and functional status. After multivariable adjustment for both patient-level and hospital-level characteristics, compared with non-Hispanic white patients, non-Hispanic black and Hispanic patients were more likely to have a longer hospital stay (i.e., length of stay greater than 4 days; black: AOR, 1.30; 95% CI, 1.22-1.39; $P<0.001$; Hispanic: AOR, 1.12; 95% CI, 1.02-1.23; $P=0.02$) and less likely to be ambulatory independent at discharge (black: AOR, 0.81; 95% CI, 0.76-0.87; $P<0.001$; Hispanic: AOR, 0.89; 95% CI, 0.79-0.99; $P=0.04$). Furthermore, non-Hispanic black patients were less likely to have a death at admission (black: AOR, 0.78; 95% CI, 0.68 to 0.89; $P<0.001$) or to be discharged home (black: AOR, 0.87; 95%CI, 0.81-0.94; $P<0.001$). Data from the Paul Coverdell National Acute Stroke Registry (PCNASR) found that other races were somewhat more likely to be assessed for rehabilitation services (98.9%) when compared to non-Hispanic white patients (98.1%) (Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, 2014).

• Bhandari VK, Kushel M, Price L, Schillinger D. Racial disparities in outcomes of inpatient stroke rehabilitation. Arch Phys Med Rehabil. 2005;86:2081-2086.
• Brown AF, Liang LJ, Vassar SD, Stein-Merkin S, Longstreth WT Jr, Ovbiagele B, Yan T, Escarce JJ. Department of Neurology, UCLA GIM&HSR. Neighborhood disadvantage and ischemic stroke: the Cardiovascular Health Study (CHS). Stroke. 2011;42(12): 3363-3368.
• Centers for Disease Control and Prevention (CDC). Outpatient rehabilitation among stroke survivors: 21 staes and the District of Columbia, 2005. MMWR. 2007;56:504-507.
• Centers for Disease Control and Prevention - Division for Heart Disease and Stroke Prevention, Annual Report, 2014.
• Chiou-Tan FY, Keng MJ Jr, Graves DE, Chan KT, Rintala DH. Racial/ethnic differences in FIM scores and length of stay for underinsured patients undergoing stroke inpatient rehabilitation. Am J Phys Med Rehabil. 2006;86:415-423.
• Chiu M, Austin PC, Manuel DG, Tu JV. Camparison of cardiovascular risk profiles among ethnic groups using population health surveys between 1996 and 2007. CMAJ. 2010;182(8):E301-10.

• Cook C, Stickley L, Ramey K, Knotts VJ. A variables associated with occupational and physical therapy stroke rehabilitation utilization and outcomes. J Allied health. 2005;34:3-10.
• Cruz-Flores S, Rabinstein A, Biller J, Elkind MSV, Griffith P, Gorelick PB, Howard G, Leira EC, Morgenstern LB, Ovbiagele B, Peterson E, Rosamond W, Trimble B, Valderrama AL, on behalf of the American Heart Association Stroke Council, Concil on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Quality of Care and Outcomes Research. Racial-ethnic disparities in stroke care: the American experience. Stroke. 2011;42:2091-2116.
• Gillum RF, Kwagyan J, Obisesan TO. Division of Geriatrics, Howard University College of Medicine, Washington, DC, USA. Ethnic and geographic variation in stroke mortality trends. Stroke. 2011;42(11):3294-3296.
• Gregory PC, Han E, Morozova O, Kuhlemeier KV. Do racial disparities exist in access to inpatient stroke rehabilitation in the state of Maryland? Am J Phys Med Rehabil.2006;85:814-819.
• Horner RD, Hoenig H, Sloane R, Rubenstein LV, Kahn KL. Racial differences in the utilization of inpatient rehabilitation services among elderly stroke patients. Stroke. 1997;28:19-25.
• Horner RD, Swanson JW, Bosworth HB, Matchar DB. VA Acute Stroke (VAST) Study Team. Effects of race and poverty on the process and outcome of inpatients rehabilitation services among stroke patients. Stroke. 2003;34:1027-1031.
• Howard G, Cushman M, Kissela BM, Kleindorfer DO, McClure LA, Safford MM, Rhodes JD, Soliman EZ, Moy CS, Judd SE, Howard VJ; REsons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Traditional risk factors as the underlying cause of racial disparities in stroke: lessons from the half-full (empty?) glass. Stroke. 2011;(12):3369-3375.
• Howard G, Howard VJ; Reasons for Geographic and Racial Differences in Stroke (REGARDS) Investigators. Ethnic disparities in stroke: the scope of the problem. Ethn Dis. 2001;11:761-768.
• Jacobs BS, Birbeck G, Mullard AJ, Hickenbottom S, Kothari R, Roberts S, Reeves MJ. Quality of hospital care in African Americans and white patients with ischemic stroke and TIA. Neurology.2006;66:809-814.
• Johnston SC, Fung LH, Gillum LA, Smith WS, Brass LM, Lichtman JH, Brown AN. Utilization of intravenous tissue-type plasminogen activator for ischemic stroke at academic medical centers.: the influence of ethnicity. Stroke. 2001; 32:1061-1068.
• Karve S, Balkrishnan R, Seiber E, Nahata M, Levine DA. Department of health Economics, RTI Health Solutions, Research Triangle Park, North Carolina. Population trends and disparities in outpatient utilization of neurologists for ischemic stroke. J Stroke Cerebrovasc Dis. 2011.
• Kissela BM, Khoury JC, Alwell K, Moomaw CJ, Woo D, Adeoye O, Flaherty ML, Khatri P, Ferioli S, De Los Rios La Rosa F, Broderick JP, Kleindorfer DO. Age at stroke: temporal trends in stroke incidence in a large, biracial population. Neurology. 2012;79:1781–1787.
• Kleindorfer DO, Khoury J, Moomaw CJ, Alwell K, Woo D, Flaherty ML, Khatri P, Adeoye O, Ferioli S, Broderick JP, Kissela BM. stroke incidence is decreasing in whites but not in blacks: a population-based estimate of temporal trends in stroke incidence from the Greater Cincinnati/Northern Kentucky Stroke Study. Stroke. 2010;41:1326–1331.
• Mayer-Oakes SA, Hoenig H, Atchison KA, Lubben JE, De Jong F, Schweitzer SO. Patient-related predictors of rehabilitation use for community-dwelling older Americans. J AM Geriatr Soc. 1992;40:336-342.
• Morgenstern LB, Smith MA, Sánchez BN, Brown DL, Zahuranec DB, Garcia N, Kerber KA, Skolarus LE, Meurer WJ, Burke JF, Adelman EE, Baek J, Lisabeth LD. Persistent ischemic stroke disparities despite declining incidence in Mexican Americans. Ann Neurol. 2013;74:778–785.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• Ottenbacher KJ, Campbell J, Kuo YF, Deutsch A, Ostir GV, Granger CV. Racial and ethnic differences in postacute rehabilitation outcomes after stroke in the United States. Stroke. 2008;39:1514-1519.
• Qian F, Fonarow GC, Smith EE, et al. Racial and ethnic differences in outcomes in older patients with acute ischemic stroke. Circ Cardiovasc Qual Outcomes. 2013;6: 284-292.
• Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS, Fullerton HJ, Gillespie C, Hailpern SM, Heit JA, Howard VJ, Kissela BM, Kittner SJ, Lackland DT, Lichtman, JH, Lisabeth LD, Makuc DM, Marcus GM, Marelli A, Matchar DB, Moy CS, Mozaffarian D, Mussolino ME, Nichol G, Paynter NP, Soliman EZ, Sorlie PD, Sotoodehnia N, Turan TN, Virani SS, Wong ND, Woo D, and Turner MB. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation. 2012;125:e78-e82.
• Stineman MG, Ross RN, Hamilton BB, Maislin G, Bates B, Granger CV, Asch DA. Inpatient rehabilitation after stroke: a comparison of lengths of stay and outcomes in the Veterans Affairs and non Veterans Affairs health care system. Med Care. 2001;39:123-137.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR

- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare
Affects large numbers, Patient/societal consequences of poor quality, Severity of illness
1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.
Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. From 2001 to 2011, the relative rate of stroke death fell by 35.1% and the actual number of stroke deaths declined by 21.2%.; however, one of every 20 deaths in the United States remains attributable to stroke. More women than men die of stroke each year. Women accounted for almost 60% of US stroke deaths in 2008 (Mozaffarian D, et al., 2015).

Stroke is also a leading cause of long-term disability (George M, et al., 2009). Stroke was among the top 18 diseases contributing to years lived with disability in 2010; of these 18 causes, only the age-standardized rates for stroke increased significantly between 1990 and 2010 ($P<0.05$)(US Burden of Disease Collaborators, 2013). Among Medicare patients discharged from the hospital after stroke, ˜45% return directly home, 24% are discharged to inpatient rehabilitation facilities, and 31% are discharged to skilled nursing facilities. Of stroke patients returning directly home, 32% use home healthcare services. In 2011, the direct and indirect cost of stroke was $33.6 billion (Mozaffarian D, et al., 2015).

The evidence indicates that patients do better with a well-organized, multidisciplinary approach to post-acute rehabilitation after stroke. The rehabilitation team may consist of a physician, nurse, physical therapist, occupational therapist, kinesiotherapist, speech and language pathologist, psychologist, recreational therapist, patient, and family members/caregivers. If an organized rehabilitation team is not available in the facility, the evidence indicates that patients with moderate or severe symptoms should be offered a referral to a facility with such a service.

Stroke rehabilitation should begin during the acute hospitalization, as soon as the diagnosis of stroke is established and life-threatening problems are controlled. The highest priorities of early stroke rehabilitation are to prevent recurrence of stroke, manage comorbidities, and prevent complications related to immobility, dysphagia, and bowel and bladder dysfunction. Rehabilitation services may include: dysphagia treatment and management; speech therapy for communication disorders (i.e., aphasia and dysarthria) and related cognitive impairments; lower-extremity strengthening and gait training; positioning, passive stretching, range-of-motion exercises, and pharmacotherapy for patients with paretic limbs and muscle spasticity; corrective measures (e.g., splinting, serial casting, surgery) for contractures; treatment interventions for post-stroke shoulder pain; treatment for depression and other cognitive and emotional disorders; and, other services. Recently published literature emphasizes the importance of identifying and managing cognitive-communication impairments after stroke since these impairments are associated with higher rate of death and higher risk of future stroke (Hinckley, 2014).

Living with disabilities after a stroke is a lifelong challenge during which people continue to seek and find ways to compensate for or adapt to persisting neurological deficits. For many, the real work of recovery begins after formal rehabilitation when the patient attempts to use newly learned skills without the support of the rehabilitation environment or team (Bates, 2005).

1c.4. Citations for data demonstrating high priority provided in 1a.3
• American Academy of Physical Medicine and Rehabilitation. Rehabilitation Helps Stroke Patients Recover Skills. AAPM&R Chicago, IL Office: Author.
• American Academy of Physical Medicine and Rehabilitation. Urgency Key But Perseverance Pays Off. AAPM&R Chicago, IL Office: Author.
• Bates B, Choi JY, Duncan PW, Glasberg JJ, Graham GD, Katz RC, Lamberty K, Recker D, Zorowitz R. American Heart Association/American Stroke Association-endorsed practice guideline. Veterans Affairs/Department of Defense clinical practice guideline for the management of adult stroke rehabilitation care. Stroke. 2005;36:2049.
• George M, Xin T, McGruder H, Yoon P, Rosamond W, Winquist A, Hinchey J, Wall H, Pandey D. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults-United States 2005. MMWR. 2009;58:421-26.
• Hinckley JJ. A case for the implantation of cognitive-communication screenings in acute stroke. American Journal of Speech-Language Pathology. 2014;23:4-14.

• Management of patients with stroke. Rehabilitation, prevention and management of complications, and discharge planning, Scottish Intercollegiate network Guidelines Network (SIGN), 2002.
• Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.
• National Institute of Neurological Disorders. Post-Stroke Rehabilitation Fact Sheet. National Institute of Neurological Disorders Bethesda, MD Office: Author.
• Post Stroke Rehabilitation, Clinical Practice Guideline No.16, Agency for Health Care Policy and Research (now known as Agency for Healthcare Research and Quality), 1995.
• VA/DoD Clinical Practice Guideline for the Management of Stroke Rehabilitation in the Primary Care Setting, 2010.
• Zorowitz RD, et al. the Post-Stroke Rehabilitation Outcomes Project (PSROP), Top Stroke Rehabil. 2005 Fall;12(4).

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Functional Status

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Appendix_A.1-635883794720981362.xls

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Since the last endorsement date, there have been no significant changes to STK-10 Assessed for Rehabilitation. Minor clarifications regarding the composition of the rehabilitation team have been made, and advanced practice nurse (APN) and physician assistant (PA) added to physician documentation as acceptable data sources for a rehabilitation assessment. Clarification was needed to prevent inclusion of other types of clinical assessments, e.g. activities of daily living, fall risk, etc., in the numerator population.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Ischemic or hemorrhagic stroke patients assessed for or who received rehabilitation services.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
One data element is used to calculate the numerator:
• Assessed for Rehabilitation Services – Documentation that the patient was assessed for or received rehabilitation services during this hospitalization. Allowable values: Yes, No/UTD or unable to determine from medical record documentation.

Patients are eligible for the numerator population when the allowable value equals “yes” for the data element.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Ischemic or hemorrhagic stroke patients.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Eight data elements are used to calculate the denominator:

1. Admission Date – The month, day and year of admission to acute inpatient care.
2. Birthdate - The month, day and year the patient was born.
3. Clinical Trial - Documentation that during this hospital stay the patient was enrolled in a clinical trial in which patients with stroke were being studied. Allowable values: Yes or No/UTD.
4. Comfort Measures Only – The earliest day the physician/APN/PA documented comfort measures only after hospital arrival. Allowable values: 1 (Day 0 or 1); 2 (Day 2 or after); 3 (Timing Unclear); 4 (Not Documented/UTD).
5. Discharge Date – The month day and year the patient was discharged from acute care, left against medical advice or expired during the stay.
6. Discharge Disposition – The place or setting to which the patient was discharged on the day of hospital discharge.
7. Elective Carotid Intervention – Documentation demonstrates that the current admission is solely for the performance of an elective carotid intervention (e.g., elective carotid endarterectomy, angioplasty, carotid stenting).
Allowable values: Yes or No/UTD.
8. ICD-10-CM Principal Diagnosis Code - The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.

Population: Discharges with an ICD-10-CM Principal Diagnosis Code for ischemic or hemorrhagic stroke as defined in Appendix A, Table 8.1 or Table 8.2.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
- Less than 18 years of age
- Length of Stay > 120 days
- Comfort measures only documented
- Enrolled in clinical trials related to stroke
- Admitted for elective carotid intervention
- Discharged to another hospital
- Left against medical advice
- Expired
- Discharged to home for hospice care
- Discharged to a health care facility for hospice care

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
- The patient age in years is equal to the Discharge Date minus the Birthdate. Patients less than 18 years are excluded.
- The Length of Stay (LOS) in days is equal to the Discharge Date minus the Admission Date. If the LOS is greater than 120 days, the patient is excluded.
- Patients with Comfort Measures Only allowable value of 1 (Day 0 or 1), 2 (Day 2 or after), and 3 (Timing unclear) are excluded.
- Patients are excluded if "Yes" is selected for Clinical Trial.
- Patients with ICD-10-PCS procedure codes for carotid intervention procedures as identified in Appendix A, Table 8.3, if medical record documentation states that the patient was admitted for the elective performance of this procedure.
- Patients with Discharge Disposition allowable value of 2 (Hospice-Home), 3 (Hospice-Health Care Facility), 4 (Acute Care Facility), 6 (Expired), or 7 (Left Against Medical Advice/AMA) are excluded.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1*

page should be provided in an Excel or csv file in required format with at S.2b)
Not applicable, the measure is not stratified.

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
Not applicable

S.16. Type of score:
Rate/proportion
If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. Calculation Algorithm/Measure Logic *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
1. Start processing. Run cases that are included in the Stroke (STK) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check Discharge Disposition
a. If Discharge Disposition equals 2, 3, 4, 6, 7, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
b. If Discharge Disposition equals 1, 5, 8, continue processing and proceed to Comfort Measures Only.

3. Check Comfort Measures Only
a. If Comfort Measures Only is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Comfort Measures Only equals 1, 2, or 3, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Comfort Measures Only equals 4, continue processing and proceed to Clinical Trial.

4. Check Clinical Trial
a. If Clinical Trial is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Clinical Trial equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population. Stop processing.
c. If Clinical Trial equals No, continue processing and proceed to Elective Carotid Intervention.

5. Check admitted for Elective Carotid Intervention
a. If Elective Carotid Intervention is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Elective Carotid Intervention equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the

Measure Population. Stop processing.
c. If Elective Carotid Intervention equals No, continue processing and proceed to Assessed for Rehabilitation Services.

6. Check Assessed for Rehabilitation Services
a. If Assessed for Rehabilitation Services is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Assessed for Rehabilitation Services equals No, the case will proceed to a Measure Category Assignment of D and will be in the Measure Population. Stop processing.
c. If Assessed for Rehabilitation Services equals Yes, the case will proceed to a Measure Category Assignment of E and will be in the Numerator Population. Stop processing.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter for the measure set cannot sample.

Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

Quarterly Sampling
The Quarterly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing quarterly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If "N" >/= 900, then "n" 180
If "N" 226-899, then "n" 20% of Initial Patient Population size
If "N" 45-225, then "n" 45
If "N" 6-44, No sampling; 100% Initial Patient Population required
If "N" 0-5, Submission of patient level data is not required; if submission occurs, 100% Initial Patient Population required

Monthly Sampling
The Monthly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for the STK Measure Set. Hospitals performing monthly sampling for STK must ensure that their Initial Patient Population and sample sizes meet the following conditions:
If "N >/= 300, then "n" 60
If "N 76-299, then "n" 20% of Initial Patient Population size
If "N 15-75, then "n" 15
If "N < 15, No sampling; 100% Initial Patient Population required

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. This measure is not based on a survey or a PRO-PM.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.24.
Electronic Clinical Data, Paper Medical Records

S.24. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Each data element in the data dictionary includes suggested data sources. The data are collected using contracted Performance Measurement Systems (vendors) that develop data collection tools based on the measure specifications. The tools are verified and tested by Joint Commission staff to confirm the accuracy and conformance of the data collection tool with the measure specifications. The vendor may not offer the measure set to hospitals until verification has been passed.

S.25. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0441_MeasureTesting_MSF5.0_Data_-1-.doc

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 0441 NQF Project: Neurology Project

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria)
Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on this measure are as follows:
170 health care organizations representing various hospital demographics:
17 For Profit; 144 Not for Profit; 9 Government
62 >=300 beds; 71 100-299 beds; 37 <100 beds
142 Urban; 28 Rural
22 Teaching; 148 Non-teaching

States represented in this data collection effort include: AL, AR, AZ, CA, CT, FL, GA, IA, ID, IL, IN, KY, LA, MA, MD, MI, MN, MO, MS, NC, NE, NJ, NM, NY, OH, PA, PR, SC, SD, TN, TX, VA, WA, WI

15 performance measurement systems are used for data transmission to The Joint Commission.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:
At the time this measure was originally tested, extensive tests of measure reliability were conducted. Pilot testing of this measure was conducted in 2004-2005 and consisted of a twelve month data collection period using a retrospective approach with monthly data transmission to The Joint Commission. To assess reliability, data were re-abstracted retrospectively by Joint Commission staff from a randomly selected sample of approximately 900 patient records identified at 30 pilot sites over the 12 month period. Results of re-abstraction were compared with original abstraction results in order to determine the rates of agreement. The objectives of pilot testing were to evaluate reliability of individual data elements, assess data collection effort and identify potential measure enhancements.

Currently, hospitals are supported in their data collection and reporting efforts by 15 contracted performance measurement system (PMS) vendors. It is a contractual requirement of Joint Commission listed vendors that the quality and reliability of data submitted to them by contracted health care organizations must be monitored on a quarterly basis. In addition, The Joint Commission analyzes these data by running 17 quality tests on the data submitted into ORYX. (ORYX is the term used by The Joint Commission to describe the component of the hospital accreditation program which requires data collection and reporting on standardized national performance measures.) The following is a list of the major tests conducted on the submitted data for this measure:

- Transmission of complete data
- Usage of data received (to understand if the healthcare organization provides the relevant service to treat the relevant population)
- Investigation of aberrant data points
- Verification of patient population and sample size

- Identification of missing data elements
- Validation of the accuracy of target outliers
- Data integrity
- Data corrections

Data Element Agreement Rate:
Inter-rater reliability testing methodology utilized by contracted performance measure system vendors is as follows:
- All clinical data elements and all editable demographic elements are scored.
- All measure data are reabstracted with originally abstracted data having been blinded so that the reabstraction is not biased.
- Reabstracted data are compared with originally abstracted data on a data element by data element basis. A data element agreement rate is calculated. Clinical and demographic data are scored separately, and an overall agreement rate is computed.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:
Data element agreement rates reported to The Joint Commission for the time period of one (4Q2010 – 3Q2011) year have shown an overall agreement rate of 98.3%. This reflects the findings of 77 hospitals, comprising 739 records. The following table delineates calculated agreement rates for individual data elements that are used to compute measure rates for STK-10.

Data Elements	Total 'n' numerator	Total 'n' denominator	Rate
Assessed for Rehabilitation Services	513	528	97.2%
Clinical Trial	712	714	99.7%
Comfort Measures Only	709	714	99.3%
Elective Carotid Intervention	711	714	99.6%

These agreement rates are considered to be well within acceptable levels.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* are consistent with the evidence cited in support of the measure focus (*criterion 1c)* and identify any differences from the evidence:
This measure focuses on determining the proportion of ischemic or hemorrhagic stroke patients who were assessed for or who received rehabilitation services. Forty percent of stroke patients are left with moderate functional impairments and 15% to 30% with severe disability (Bates B, et al., 2005). These measure specifications are consistent with clinical practice guidelines that recommend for an initial assessment of complications, impairment and rehabilitation needs, in addition to obtaining the medical history and physical examination, and assessment of stroke severity. Post acute stroke patients whose assessment determines the need for rehabilitation services should be offered organized and coordinated inpatient stroke rehabilitation care provided by a multi-disciplinary team of rehabilitation professionals. The measure specifications also address this aspect of stroke rehabilitation.

Measure specifications differ from clinical practice guidelines in that the components of the rehabilitation assessment are not included in the specifications. Additionally, patients and family members with severe disability and a poor prognosis for functional recovery who are assessed to be ineligible for rehabilitation services should receive counseling and advice regarding future plans for medical follow-up. This measure does not address the educational component needed for severe stroke patients.

Patients less than 18 years of age that have a length of stay (LOS) more than 120 days, enrolled in a clinical trial for stroke or who were designated "comfort measures only" anytime during hospitalization are excluded to harmonize with other CMS/Joint Commission measures. Operationally, patients with a planned admission to the hospital for an elective carotid intervention which restores blood flow to the brain and prevents stroke, such as carotid endarterectomy or carotid stenting, are excluded from the measure.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:

At the time this measure was originally tested measure validity was assessed via survey and focus groups of hospitals participating in the pilot test. All measure specifications, including population identification, numerator and denominator statements and exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant.

Since the measure has been in national use, continued validity of the measure has been determined through analysis of feedback from measure users. The Joint Commission provides a web-based application with which measure users can provide feedback regarding appropriateness of measure specifications, request clarification of specifications, and/or provide other comments pertinent to the measure. This feedback is systematically, continually, reviewed in order to identify trends and to identify areas of the measure specifications that require clarification or revision. An expert technical advisory panel (TAP) meets on a quarterly basis in order to assess continued validity of this measure. Additionally, Joint Commission staff continually monitors the national literature and environment in order to assess continued validity of this measure. And finally, the conversion from ICD-9-CM diagnosis codes to ICD-10-CM diagnosis codes has been completed and reviewed by the Technical Advisory Panel for validity. The panel has determined that the intent of the measure has not changed as a result of the conversion. The crosswalk will also be posted in a future version of the specifications manual for public comment, and results of feedback will be reviewed and incorporated into the measure specifications where indicated.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:
Analysis of feedback obtained via our automated feedback system reveals 72 submissions regarding specifications for this measure over the past year. Predominantly, questions involve the data element Assessed for Rehabilitation Services. Questions generally pertain to the composition of the rehabilitation team and the qualifications needed to complete the rehabilitation assessment despite detailed inclusion guidelines.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
As noted previously, this measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:
Measure exclusions that were not derived directly from the evidence are presented below. Please note that these are population exclusions that are necessary to ensure consistency in all measures in this 8 measure set.
These exclusions were analyzed for frequency of occurrence. An issue that is of great concern to users of this measure is that due to the presence of exceptions to the measure, attainment of a 100% measure rate is not possible. Because of the role of this measure in the current Joint Commission accreditation process this is especially troubling to measure users. This concern is the basis for a number of the non-evidence-based exclusions to these measures. The following measure exclusions that were not derived directly from the evidence are as follows:

1. Patients < 18 years old
2. Patients who have a length of stay (LOS) greater than 120 days
3. Patients with Comfort Measures Only documented
4. Patients enrolled in clinical trials
5. Patients admitted for Elective Carotid Intervention
6. Patients discharged to another hospital (acute care facility)
7. Patients who left against medical advice
8. Patients who expired
9. Patients discharged to home for hospice care
10. Patients discharged to a health care facility for hospice care

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*
N=39,812

1. Patients < 18 years old = 0%
2. Patients who have a length of stay (LOS) greater than 120 days = 0%

3. Patients with Comfort Measures Only documented = 11.11%
4. Patients enrolled in clinical trials = 0.39%
5. Patients admitted for Elective Carotid Intervention = 9.96%
6. Patients discharged to another hospital (acute care facility) = 1.13%
7. Patients who left against medical advice = 0.25%
8. Patients who expired = 3.58%
9. Patients discharged to home for hospice care = 0.60%
10. Patients discharged to a health care facility for hospice care = 1.52%

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Not Applicable

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:
Not Applicable

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:
Not Applicable

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:
Not Applicable

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
This measure has been in national use since the 4th quarter of 2009. Demographics of organizations collecting and reporting data on these measures are as previously reported.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:
The method used to analyze meaningful differences in performance at The Joint Commission is Target Analysis. The object of target analysis is to compare a health care organization's data against a comparative norm for the purpose of evaluating performance improvement opportunities. When an organization's performance level is statistically significantly different from a comparative norm, it is considered a statistical deviation. A statistical deviation may be desirable or undesirable depending on the "direction of improvement" of the measure.

There are two components to the target analysis methodology used at The Joint Commission. Given the national average for a performance measure, a target range is constructed. Using generalized linear mixed models methodology (also known as hierarchical models), a predicted estimate of a HCO's performance, with a corresponding 95% confidence interval, is generated. This confidence interval is compared to the target range, to determine the HCO's rating. The estimate of the organization's true performance is based on both the data from that organization and on data from the entire set of reporting organizations.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:
STK-10 Distribution of Measure Results

Quarter	#hospitals	Mean	SD	90th%	75th%	50th%	25th%	10th Percentile
3Q2011	154	0.95859	0.1098	1	1	1	0.95833	0.88889
2Q2011	155	0.95968	0.09234	1	1	1	0.95918	0.9

1Q2011	160	0.95387	0.09443	1	1	1	0.94595	0.85165
4Q2010	136	0.95043	0.11751	1	1	1	0.95339	0.85714
3Q2010	131	0.95645	0.07732	1	1	0.99038	0.95	0.87059
2Q2010	122	0.95388	0.0842	1	1	0.99403	0.94444	0.88119
1Q2010	99	0.94899	0.0808	1	1	0.98333	0.93421	0.84615
4Q2009	49	0.95212	0.07659	1	1	0.97744	0.95	0.85294

It should be noted that since data collection on this measure is completely voluntary for both The Joint Commission and PCNASR, the hospitals reporting on this measure are self-selected, and therefore, presumably have a particular interest and concern for improving stroke care. For this reason, the measure results demonstrated by this group most likely significantly overstates the rate for the population of all health care organizations.

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Multiple data sources are not used for this measure.

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:
Not applicable

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:
Not applicable

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: The measure is not stratified for disparities.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:
Few studies have examined racial-ethnic disparities in stroke rehabilitation, and the evidence available is somewhat unclear and contradictory. Hence, there are no plans to stratify the measure. The Joint Commission does not currently capture gender-specific data, or data elements for race or ethnicity because these data elements have not been shown to be reliably collectable due to the fact that no national standardized definitions exist for these data elements. Also, not all hospitals collect race and ethnicity. In the future, it may be feasible for The Joint Commission to explore how race and ethnicity and other relevant disparity data, might be collected reliably.

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?
(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

NEW (since 2012 endorsement): Data for Empirical Testing

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

2b2. VALIDITY TESTING
2b2.1. What level of validity testing was conducted? (*may be one or both levels*)
☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The data used to measure validity consists of one year data (third, and fourth quarter 2014 and first and second quarter for 2015) extracted from The Joint Commission warehouse. The hospitals selection was based to those hospitals that reported 12 months of data and had 30 or more denominator cases for the year.

Measure convergent and criterion validity was assessed using hospitals patient level data from The Joint commission warehouse. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous face validity testing. The patient population was comprised of all patients discharged with an ICD-9-CM Principal Diagnosis Code for stroke as defined in Appendix A, Table 8.1 or Table 8.2.

Conversion of ICD-9 to ICD-10 equivalent codes was consistent with the clinical intent of the original measure specifications. The Joint Commission worked with a certified coding expert throughout the conversion process. The 3M Coding Conversion Tool was utilized, including forward mapping of ICD-9 codes to ICD-10 codes as well as reverse mapping from ICD-10 to ICD-9 to ensure appropriateness. MSDRGs and instructions in the tabular index were also examined to ensure appropriate code mapping. Crosswalks comprising ICD-9 codes mapped to ICD-10 codes were created and reviewed by members of the Technical Advisory Panel, CMS subcontractors, and performance measurement system vendors prior to being posted for a 12 month public comment period. Feedback from the field indicated that the crosswalks generally were mapped correctly. Minor modifications to the code tables were made as needed. Final code tables were published in early 2015, well in advance of the mandated date of October 1, 2015.

We conducted a secondary analysis to help interpret results; correlation with other stroke process measures.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Descriptive statistics for the STK measures:
N=1,318 hospitals
n= 2, 206,379 patients records

STK-10
Median: 100%
Percentile 10%: 95%
Percentile 25%: 98%
Percentile 75%: 100%
Percentile 90%: 100%

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations							
	STK_1	STK_2	STK_3	STK_4	STK_5	STK_6	STK_10
STK_1	1.00000 1318	0.55423 <.0001 1318	0.30311 <.0001 1302	0.37009 <.0001 1144	0.52785 <.0001 1318	0.53082 <.0001 1318	0.45291 <.0001 1318
STK_2	0.55423 <.0001 1318	1.00000 1318	0.37566 <.0001 1302	0.28169 <.0001 1144	0.53355 <.0001 1318	0.47326 <.0001 1318	0.29578 <.0001 1318
STK_3	0.30311 <.0001 1302	0.37566 <.0001 1302	1.00000 1302	0.25551 <.0001 1137	0.25665 <.0001 1302	0.28198 <.0001 1302	0.15819 <.0001 1302
STK_4	0.37009 <.0001 1144	0.28169 <.0001 1144	0.25551 <.0001 1137	1.00000 1144	0.22516 <.0001 1144	0.43717 <.0001 1144	0.41076 <.0001 1144
STK_5	0.52785 <.0001 1318	0.53355 <.0001 1318	0.25665 <.0001 1302	0.22516 <.0001 1144	1.00000 1318	0.35302 <.0001 1318	0.39079 <.0001 1318
STK_6	0.53082 <.0001 1318	0.47326 <.0001 1318	0.28198 <.0001 1302	0.43717 <.0001 1144	0.35302 <.0001 1318	1.00000 1318	0.45420 <.0001 1318
STK_10	0.45291 <.0001 1318	0.29578 <.0001 1318	0.15819 <.0001 1302	0.41076 <.0001 1144	0.39079 <.0001 1318	0.45420 <.0001 1318	1.00000 1318

The correlations of the stroke measures are all positively correlated and statistically significant indicating that hospitals with high quality on one stroke measure tend to have high performance on the other stroke measures.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)
Overall, the positive inter-correlations indicates convergent validity of the measures. STK-10 is positively correlated with other stroke evidence-based processes of care.

2b3. EXCLUSIONS ANALYSIS .
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patient who have a Length of Stay greater than 120 days
- Patients with *Comfort Measures*

- Patient enrolled in a Clinical Trial
- Patients admitted for *Elective Carotid Intervention*
- Patients discharged to another hospital
- Patients who left against medical advice
- Patients who expired
- Patients discharged to home for hospice care
- Patients discharged to a health care facility for hospice care

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2, 206,379 admissions included in the initial cohort. From among the 2, 206,379 admissions in 1,318 hospitals, the descriptive statistics are given below.

Exclusion: *Comfort Measures Only*
Overall Occurrence n = 163,225
Overall Occurrence Percentage: 10.4%
Minimum: 0.30%
10th Percentile: 5.26%
Median: 10%
90th Percentile: 15.8%
Maximum: 35.7%

Exclusion: Clinical Trial
Overall Occurrence n = 5,446
Overall Occurrence Percentage: 0.25%
Minimum: 0.08%
10th Percentile: 0.24%
Median: 0.52%
90th Percentile: 1.90%
Maximum: 10%

Exclusion: Patients admitted for *Elective Carotid Intervention*
Overall Occurrence n = 259,833
Overall Occurrence Percentage: 11.8%
Minimum: 0.16%
10th Percentile: 2.28%
Median: 11.5%
90th Percentile: 25.3%
Maximum: 95.2%

Exclusion: *Discharge Disposition* - Patients discharged to another hospital
Overall Occurrence n = 449,924
Overall Occurrence Percentage: 35.7%
Minimum: 0.787%
10th Percentile: 25%
Median: 35.4%
90th Percentile: 46%
Maximum: 76.2%

Exclusion: *Discharge Disposition* - Patients who left against medical advice
Overall Occurrence n = 8,396
Overall Occurrence Percentage: 0.67%
Minimum: 0.067%

10th Percentile: 0.34%
Median: 0.86%
90th Percentile: 2.4%
Maximum: 9.67%

Exclusion: *Discharge Disposition* - Patients who expired
Overall Occurrence n = 76,168
Overall Occurrence Percentage: 6.04%
Minimum: 0.398%
10th Percentile: 1.9%
Median: 5.08%
90th Percentile: 10.2%
Maximum: 20.1%

Exclusion: *Discharge Disposition* - Patients discharged to home for hospice care
Overall Occurrence n = 658,264
Overall Occurrence Percentage: 52.2%
Minimum: 6.25%
10th Percentile: 39%
Median: 51.9%
90th Percentile: 64%
Maximum: 94.3%

Exclusion: *Discharge Disposition* - Patients discharged to a health care facility for hospice care
Overall Occurrence n = 37,804
Overall Occurrence Percentage: 3%
Minimum: 0.169%
10th Percentile: 0.86%
Median: 3.01%
90th Percentile: 6.4%
Maximum: 23%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The median frequency of exclusions range from low to moderate. The distribution of exclusions across hospitals is generally narrow indicating that the occurrence is random and likely would not bias performance results, although the percentage of patients excluded may differ depending on whether the hospital was a stroke center.

For criterion validity, it is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a Length of Stay greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Patients with *Comfort Measures Only* documented
Rationale: It is inappropriate to treat such patients beyond those measures necessary for comfort.

Patients enrolled in a Clinical Trial
Rationale: Rehabilitation services as specified by this measure may compromise a clinical trial.

Patients admitted for *Elective Carotid Intervention*
Rationale: Patients who are not admitted for treatment of an acute stroke may not require rehabilitation services

Patients discharged to another hospital
Rationale: This measure is meant for patients discharged to home.

Patients who left against medical advice
Rationale: Hospitals do not have opportunity for provision of quality care for the non-compliant patient.

Patients who expired
Rationale: Patients who expire are not eligible to be in this measure

Patients discharged to home for hospice care
Rationale: Rehabilitation services may not be warranted for the hospice patient

Patients discharged to a healthcare facility for hospice care
Rationale: Rehabilitation services may not be warranted for the hospice patient.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other

If other: Data element allowable values are selected, either manually or electronically, from clinical and coded data available in medical record documentation. All medical record documentation is used in the abstraction process. Vendor data collection tools are used to import data elements needed for measure rate calculation.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract the electronic version of this measure, so continues to offer this chart-abstracted version which allows for data capture from unstructured data fields. All data elements needed to compute the STK-10 performance measure score have been retooled for capture from electronic sources. Annual updates are performed to match the eCQM specifications to the current version of the chart-abstracted specifications.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

At the present time, hospitals using this performance measure generally collect measure data via manual review of the paper medical record, the EHR or a combination of both. Collected data are submitted to The Joint Commission on a quarterly basis, by way of contracted performance measurement system vendors, as described previously. Specifications for this measure are freely available to anyone who wishes to use the measure. Feedback from hospitals using this measure indicates that required data elements are generally available in the medical record, and measure specifications are robust and easy to understand. As described above, as feedback from measure users has indicated the need for clarification or revision of measure specifications, this has taken place in the form of guidelines for abstraction. Specific revisions are detailed in the Release Notes section of this submission.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

There are no fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public

domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Quality Check® http://www.qualitycheck.org/consumer/searchQCR.aspx Hospital Compare https://www.medicare.gov/hospitalcompare/search.html Public Health/Disease Surveillance Paul Coverdell National Acute Stroke Registry http://www.cdc.gov/dhdsp/programs/stroke_registry.htm Payment Program Hospital Inpatient Quality Reporting Program https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalrhqdapu.html Regulatory and Accreditation Programs Hospital Accreditation Program http://www.jointcommission.org/ Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Annual Report – Improving America's Hospitals http://www.jointcommission.org/annualreport.aspx Quality Improvement (Internal to the specific organization) Disease-Specific Care Certification for Comprehensive Stroke Centers Disease-Specific Care Certification for Primary Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

•Name of program and sponsor: Quality Check®; The Joint Commission

•Purpose: A public website that allows consumers to: search for accredited and certified organizations by city and state, by name or by zip code (up to 250 miles); find organizations by type of service provided within a geographic area; download free hospital performance measure results; and, print a list of Joint Commission certified disease-specific care programs and health care staffing firms.

•Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

•Name of program and sponsor: Hospital Compare; Centers for Medicare & Medicaid Services

•Purpose: A public website that provides information that helps consumers decide where to obtain healthcare and encourages hospitals to improve the quality of care they provide.

•Geographic area and number and percentage of accountable entities and patients included: Nationwide; 4000+ medicare-certified hospitals (2015)

•Name of program and sponsor: Paul Coverdell National Acute Stroke Registry; Centers for Disease Control and Prevention

•Purpose: A national registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care.

•Geographic area and number and percentage of accountable entities and patients included: 11 states; 403 hospitals (CDC, 2014)

•Name of program and sponsor: Hospital Inpatient Quality Reporting Program; Centers for Medicare & Medicaid Services

•Purpose: The Hospital Inpatient Quality Reporting (Hospital IQR) program was mandated by Section 501(b) of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA) of 2003. This section of the MMA authorized CMS to pay hospitals that successfully report designated quality measures a higher annual update to their payment rates.

•Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3500 hospitals (2015)

•Name of program and sponsor: Annual Report-Improving America's Hospitals; The Joint Commission

•Purpose: The annual report summarizes the performance of Joint Commission-accredited hospitals on 46 accountability measures of evidence-based care processes closely linked to positive patient outcomes, and provides benchmarks from Top Performer on Key Quality Measures® hospitals.

•Geographic area and number and percentage of accountable entities and patients included: Nationwide; 3300 Joint Commission-accredited hospitals (2014)

•Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint Commission

•Purpose: A certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.

•Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

•Name of program and sponsor Disease-Specific Care Certification for Primary Stroke Centers; The Joint Commission

•Purpose: A certification program that recognizes hospitals that effectively manage and meet the unique and specialized needs of stroke patients, and make exceptional efforts to foster improved outcomes for better stroke care.

•Geographic area and number and percentage of accountable entities and patients included Nationwide; 1079 hospitals

•Name of program and sponsor Hospital Accreditation Program; The Joint Commission

•Purpose: An accreditation program that recognizes hospitals that meet standard requirements to provide safe and effective patient care.

•Geographic area and number and percentage of accountable entities and patients included Nationwide; 3300 Joint Commission-accredited hospitals (2014)

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

•Progress (trends in performance results, number and percentage of people receiving high-quality healthcare):

Over the past five years, there has been steady but small increases of ~1% per year in the percentage of patients assessed for rehabilitation services during the inpatient hospitalization based on Joint Commission ORYX performance measurement data. The national aggregate rate reported for 2014 was 98.7%. A modest gap of ~5% still exists for the lowest decile of hospitals. This trend is consistent with increases in the rate of assessment for rehabilitation services published by GWTG-Stroke and PCNASR.
Racial disparities for secondary outcomes, i.e., length of stay > 4 days, ambulatory at discharge, death at admission, and discharges to home, have been reported (Qian, 3013; CDC 2014). Non-Hispanic black and Hispanic patients were more likely to have longer lengths of stay and less likely to be discharged home, according to Qian and associates. These disparities were statistically significant for non-Hispanic black patients ($P<0.001$), and support the continued importance for early assessment of rehabilitation needs prior to hospital discharge.

•Geographic area and number and percentage of accountable entities and patients included:
Nationwide; 1299 hospitals; 210,075 patients (The Joint Commission, 2014)

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

Since the last endorsement date, questions from the field have reflected that other types of assessments, e.g., fall risk, activities of daily living, discharge planning, may be abstracted as rehabilitation assessments. Abstraction guidelines for the data element Assessed for Rehabilitation have been strengthened to prevent inference. Only assessments documented by a qualified member of the rehabilitation team may be accepted as a rehabilitation assessment.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0244 : Stroke and Stroke Rehabilitation: Rehabilitation Services Ordered

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
0244 : Stroke and Stroke Rehabilitation: Rehabilitation Services Ordered – no longer NQF endorsed.

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
NQF#0244 focuses on rehabilitation orders written prior to hospital discharge and not tbe rehabilitation assessment or services received by the patient.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Although both NQF#0441 STK10: Assessed for Rehabilitation and NQF#0244 target ischemic or hemorrhagic stroke patients in the acute inpatient setting, NQF#0441 is superior for two reasons. First, the numerator statement for NQF#0441 is a broader measure of quality and encompasses the total ischemic or hemorrhagic stroke inpatient population. The proportion of ischemic or hemorrhagic stroke patients who are assessed for or receive rehabilitation services during the acute inpatient hospitalization are captured in the numerator population. Patients must be assessed before services can be ordered. Rehabilitation services may be ordered but not implemented. Consequently, rehabilitation services are not received when orders are not carried out. NQF#0441 includes stroke patients who receive rehabilitation services in the numerator population. Second, NQF#0244 focuses on rehabilitation orders written prior to hospital discharge, but capture these data after hospital discharge in the outpatient setting.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0507
Measure Title: Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports
Measure Steward: American College of Radiology (ACR)
Brief Description of Measure: Percentage of final reports for carotid imaging studies (neck magnetic resonance angiography (MRA), neck computerized tomographic angiography (CTA), neck duplex ultrasound, carotid angiogram) performed that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement
Developer Rationale: There is wide variation in the use of methods for stenosis calculation, which may also lead to variation in the appropriateness of carotid intervention. Since the degree of stenosis is an important element of the decision for carotid intervention, characterization of the degree of stenosis needs to be standardized. Requiring that stenosis calculation be based on a denominator of distal internal carotid diameter or, in the case of duplex ultrasound, velocity measurements that have been correlated to angiographic stenosis calculation based on distal internal carotid diameter, makes the measure applicable to both imaging and duplex studies.

Numerator Statement: Final reports for carotid imaging studies that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement
Denominator Statement: All final reports for carotid imaging studies (neck MRA, neck CTA, neck duplex ultrasound, carotid angiogram) performed
Denominator Exclusions: No Denominator Exclusions or Denominator Exceptions

Measure Type: Process
Data Source: Administrative claims, Electronic Clinical Data : Registry
Level of Analysis: Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Oct 28, 2008 **Most Recent Endorsement Date:** Mar 06, 2013

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☐ **Yes** ☒ **No**

Evidence Summary from **prior review in 2012**

- The developer's rationale for the measures is that standardizing the method for stenosis calculation will lead to more accurate quantification of stenosis and thus more appropriate treatment.
- The evidence cited as support for this measure includes clinical practice guidelines, systematic reviews, and additional studies.
- The developer highlighted that the Society of Radiologists in Ultrasound convened a panel (2002) that recommended
  that the NASCET method of carotid stenosis measurement should be used when angiography is used to correlate the ultrasound (US) findings.
- The developer presented several potential challenges to using the NASCET criteria including failure to assess for near occlusion and measuring the still tapering bulb.
- In the prior evaluation, one Committee member noted that the NASCET technique accurately predicts stroke risk in symptomatic patients, but questioned whether documenting usage of a standardized stenosis measurement technique improves patient outcomes. Some members questioned whether results from other measurement techniques are accurate; another member clarified that different methods yield different results, making documentation of the method necessary.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☐ **The developer provided updated evidence for this measure:**

Exception to evidence – N/A

Guidance from the Evidence Algorithm

Process measure assesses performance, based on systemic reviews, etc. (Box 3) → QQC presented (Box 4) → Moderate certainty of net benefit (Box 5b)

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and voting on Evidence?*

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap.

PLEASE DISREGARD DATA INCLUDED UNDER SECTION 1B—THE DEVELOPER HAS PROVIDED UPDATED INFORMATION (SEE END OF THIS WORKSHEET). NQF WILL ASK THE DEVELOPER TO FORMALLY INCLUDE UPDATED INFORMATION IN THEIR SUBMISSION FORM PRIOR TO THE IN-PERSON MEETING.

- The developer provided two sets of PQRS data—one for all reporting physicians and one for those who had more than 10 eligible patients for 3 years (2012-2014). These data indicate that for 2014, the median performance

rate was 98.11% for all reporting physicians and 95.15% for those with more than 10 patients.

- The developers also state that "*The 2014 rate indicates that 15.9% of patients reported on did not receive optimal care*" (this is based on those physicians with more than 10 patients). The developer indicates that the reporting rate for this measure was 85.8% in 2014.
- The average performance rate reported by the developer (77.4%) reflects an overall patient percentage, not the average rate across clinicians.

Disparities - No data on disparities were provided.

Questions for the Committee:

- *Do enough clinicians report on this measure to PQRS to make allow for an evaluation of performance?*
- *Is there a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **Maintenance - no new evidence reported. Data provide more accurate assessment of need for carotid intervention and exclusion of strokes inappropriate for tPA. Does this improve patient outcome?

**Not aware of new studies that change the evidence base. The measure assesses compliance with a standardized reporting criterion for carotid stenosis. It applies directly to the process.

1b. Performance Gap

Comments: **Performance 85.8% for 10 or more patients. There is room for improvement.

No disparity data!

**There is ample evidence that many radiology reports do not comply with the standard. There is evidence that compliance has increased substantially in recent years. Continued application of the standard will likely lead to further increases. There is no data on disparities. For this measure disparity data would be difficult to interpret, since those interpreting the studies typically do not know the race of the patients and might only infer ethnicity by name.

1c. High Priority (previously referred to as High Impact)

Comments: **There is little discussion of rationale and additive value. The additive value is difficult to estimate. Many, perhaps most, of these studies are done for indications for which this measure- carotid stenosis reporting- would be of no consequence.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- Data source**:** Administrative claims and Registry
- Level of analysis: Individual clinician
- Care setting: Hospital/Acute Care Facility, Imaging Facility
- CPT codes are used when data are obtained from claims data
- Numerator: Final reports for carotid imaging studies that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement
- Numerator Definition: Direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement - includes direct angiographic stenosis calculation based on the distal lumen as the denominator for stenosis measurement OR an equivalent validated method referenced to the above method

- Denominator Statement: All final reports for carotid imaging studies (neck MRA, neck CTA, neck duplex ultrasound, carotid angiogram) performed
- No denominator exclusions or denominator exceptions

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is it likely this measure can be consistently implemented?*
- *Are numerator instructions clear?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- Prior testing was conducted at the data element level via inter-rater reliability analysis .

Describe any updates to testing: The developers conducted score-level testing of the measure using a signal-to-noise analysis.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☐ **No**

Method(s) of reliability testing

Prior testing:

- Developers conducted Inter-rater reliability testing by comparing data gathered by two trained clinical abstractors and evaluating their agreement rate. This testing used data from 3 radiology sites during calendar year 2010; 109 records were included in the testing sample.

Updated testing:

- A physician level registry or claims database was used for extracting the relevant physician level information for 2012-2014
- The numbers of physicians were 133,717 physicians
- The data collection period was 2012-2014
- Empirical validity testing at the performance measure score level was conducted via a signal-to-noise analysis using the beta-binomial model. This is an appropriate method for testing reliability.
- The testing sample included 2012-2014 data from 133,717 physicians and 2,268,250 patients; data from both claims and registry were used.
- A signal-to-noise analysis quantifies the amount of variation in performance that is due to differences between physicians (as opposed to differences due to random measurement error). Results will vary based on the amount of variation between the hospitals and the number of patients treated by each hospital. This method results in a reliability statistic that ranges from 0 to 1 for each provider. A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in provider performance. A value of 0.7 often is regarded as a minimum acceptable reliability value.

Results of reliability testing

Prior results: Updated results:

Summary of PQRS Reliability Score Stats Cumulative and by Year (2012 - 2014)

Year	*Number of Providers*	*Reliability p25*	*Reliability median*	*Reliability p75*	*Reliability mean*	*Reliability LCLM*	*Reliability UCLM*
2012	37142	0.81728	1	1	0.84524	0.84238	0.84809
2013	33493	0.63205	1	1	0.81781	0.81477	0.82085
2014	12953	0.99306	0.99722	0.99913	0.99473	0.99461	0.99484
All	83588	0.88581	1	1	0.85741	0.85561	0.85922

Guidance from the Reliability Algorithm:

Specifications are precise (Box 1)>empirical reliability testing (Box 2)>Tested with performance measure score (Box 4)>signal-to-noise method appropriate (Box 5) > testing shows high certainly (Box 6b)>recommend high

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Developer states that they systematically assessed face validity as an indicator of quality

Describe any updates to validity testing

- Developer states that since the previous submission, this measure underwent maintenance review by an expert panel

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- 2012 face validity assessment: Face validity of the measure was assessed by asking a 14-member expert panel to rate their agreement with the following statement: *The scores obtained from the measure as specified will accurately differentiate quality across providers.* Members rated their agreement using a 5-point Likert scale, where 1=Strongly Disagree; 3=Neither Disagree nor Agree; 5=Strongly Agree.
- 2015 face validity assessment: The developer stated that since the previous submission, this measure underwent maintenance review by an expert panel. No additional information was provided.

Validity testing results:

- 2012 face validity assessment: The results of the expert panel rating of the validity statement was that 85.71% of respondents either agreed or strongly agreed that this measure can accurately distinguish good and poor quality.
- 2015 face validity assessment: According to the developer, the expert panel agreed that the measure remained valid based on existing evidence.

Questions for the Committee:

- *Do you agree that the score from this measure as specified is an indicator of quality?*
-

2b3-2b7. Threats to Validity

2b3. Exclusions:

- There are no exclusions for this measure.

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:
The developer did not provide information about meaningful differences in this section. However, their reliability analysis (section 2a above) indicates that the measure can distinguish between providers and their information on opportunity for improvement (section 1b above) indicates at least some variation in performance rates across providers.

Question for the Committee:
Can this measure identify meaningful differences about quality?

2b6. Comparability of data sources/methods:

Developer states this is not applicable; however, they did provide descriptive statistics on performance stratified by data source (section 1b above). These data indicate that performance is somewhat higher when the data source is administrative claims.

2b7. Missing Data

The developer did not provide any information on missing data.

Guidance from the Validity Algorithm:
Measures specification consistent with evidence (Box1)>some threats to validity assessed (Box 2)>no empirical validity (Box 3)>Face Validity (Box 4)>can distinguished quality(Box 5)>recommend moderate

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments:
**No change - large sample - well specified elements.

**The data elements are clearly defined. Codes are provided. Steps in the logic and calculation are clear.
**Same
**Specifications are consistent with the evidence

2a2. Reliability Testing

Comments: **Last year - all 99%

**Yes. Reliability testing was adequate in scope to generalize. This measure has been widely adopted and has been reliable in the field.

**Face validity only - high values in 2012 and 2015. Concern with only face validity. I want to see if this measure improves outcomes!
**Validity was assessed for face validity by an expert panel. The panel agreed that the measure was valid. There are no tests of whether the method of carotid reporting influences patient care or outcomes.

2b2. Validity Testing

Comments: **No Exclusions
No data on missing data

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Last year - all 99%
**Yes. Reliability testing was adequate in scope to generalize. This measure has been widely adopted and has been reliable in the field.
**Face validity only - high values in 2012 and 2015. Concern with only face validity. I want to see if this measure improves outcomes.

**Validity was assessed for face validity by an expert panel. The panel agreed that the measure was valid. There are no tests of whether the method of carotid reporting influences patient care or outcomes.

**No Exclusions

**There are no exclusions. There probably should be

**None

** The analysis indicates meaningful differences in performance.

** N/A

**No data on missing data

**Does not constitute a threat to validity

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic sources and generated or collected by and used by healthcare personnel during the provision of care.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **All data are in defined e fields.

**The measure is feasible and in wide use. Readliy derived from EHR.

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- This measure has been used in the CMS Physician Quality Reporting Initiative/System 2007-2016: PQRS#195 Radiology: Stenosis Measurement in Carotid Imaging Reports.

Improvement results

PLEASE DISREGARD DATA INCLUDED UNDER SECTION 4B—THE DEVELOPER HAS PROVIDED UPDATED INFORMATION (SEE END OF THIS WORKSHEET). NQF WILL ASK THE DEVELOPER TO FORMALLY INCLUDE UPDATED INFORMATION IN THEIR SUBMISSION FORM PRIOR TO THE IN-PERSON MEETING.

- According to the developer, the patient-level performance rate based on PQRS data for physicians with more than 10 patients increased from 2012 to 2014. For 2012 the rate was 68.7%, for 2013 it was 79.0%, and lastly for 2014 the rate was 84.1%. The median clinician-level performance rate across these years was 82.86%, 90.0%, and 95.15%.

Unexpected findings (positive or negative) during implementation

- Developer did not present any information on unexpected findings

Potential harms

- Developer states there they were not aware of any unintended consequences related to this measurement

Feedback :

- This measure was reviewed by the Measure Applications Partnership (MAP) for Physician Quality Reporting System (PQRS) program in 2012. In 2012, the measure was reviewed by MAP for Medicare and Medicaid EHR Incentive Program for Eligible Professionals. In 2014 it was reviewed for Physician Compare (PQRS) and the Value-Based Payment Modifier(VBPM). MAP did not support the use of this measure in PQRS or VBPM. The rationale provided for this was that measure did not adequately address any needs of the program.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

**All fields public and accountable*

** The measure in being used. To the extent that it contributes to health quality, reporting this measure encourages* compliance.

4b. Improvement.

4c. Unintended Consequences

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **All fields public and accountable. improved outcomes?

**The measure in being used. To the extent that it contributes to health quality, reporting this measure encourages compliance. No known unintended consequences.

**Improved outcomes?

**No known unintended consequences.

Criterion 5: Related and Competing Measures

Related or competing measures

None

Harmonization

N/A

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence.

Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. (evaluation criteria)

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process-health outcome; intermediate clinical outcome-health outcome)*:

The process of standardizing the method for stenosis calculation, as indicated in the measure language, will lead to improved health outcomes such as more accurate quantification of stenoses and more appropriate treatment, based on the percentage of stenoses.

1c.2-3 Type of Evidence *(Check all that apply)*:

Clinical Practice Guideline

Selected individual studies (rather than entire body of evidence)

Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:

The consensus statement, practice guidelines, and the body of evidence encourage the use of a standardized criteria for the quantification of carotid stenosis. This measure encourages the appropriate calculation of stenoses in all patients, regardless of diagnosis, receiving any of the identified exams (ie, neck MR angiography [MRA], neck CT angiography [CTA], neck duplex ultrasound, carotid angiogram).

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*: The Society of Radiologists in Ultrasound consensus document, included the review of 30 studies.

The systematic review by Cina et al reviewed 23 studies.

The systematic review by Koelemay, et al reviewed 28 studies.

The systematic review by Rothwell (2003) reviewed two studies.

A meta-analysis by Goldstein et al, included the review of three studies.

The systematic review by Naylor et al, includes 48 publications.

The body of evidence also includes the NASCET and ECST studies, as well as a Study in JACC by Brooks et al.

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events):* The 30 studies included in the literature review for the Society of Radiologists in Ultrasound consensus document, included the review of Doppler US Thresholds and Performance in Diagnosis of internal carotid artery stenosis. The studies also included literature specifically focused on ICA stenosis, including but not limited to the following assessments: laboratories using incosistent thresholds, use of Dopper US to estimate a single degree of stenosis, assessment of how the ratio of ICA PSV at and distal to stenosis performs vs. the ICA-to-CCA ratio, PSV for predicting >70% stenosis, Doppler parameters that did not use NASCET angiography criteria.

The systematic review by Cina et al included 23 studies that met the following criteria: comparison of carotid endarterectomy with nonsurgical management, inclusion of patients with carotid stenosis ipsilateral to neurologic symptoms or signs in the carotid territory, and randomized controlled trial design.

The systematic review by Koelemay included a review of 28 studies from a literature search between 1990 and 2003, comparing CTA and intra-arterial digital subtraction angiography in patients with symptomatic carotid artery disease that presented raw data for detection of a <70% or 70% to 99% stenosis or an occlusion.

The systematic review by Rothwell randomized 3018 patients and followed them up for a mean of 73 months, remeasuring the prerandomization ESCT carotid angiograms and redefined the outcome events the same way as in NASCET. The 3018 patients were randomized and followed up for a mean of 73 months.

The meta-analysis by Goldstein et al, analyzed data from the NASCET, ECST, and VACS studies, which were all randomized. Data for specified endpoints and corresponding person-years at risk were obtained for each study. The rates of nonfatal stroke, nonfatal myocardial infarction, or death for surgically or medically treated patients in the perioperative period (30 days) and thereafter were compared (both including and excluding perioperative events) and then combined using meta-analytic techniques.

In the North American Symptomatic Carotid Endarterectomy trial, fifty North American centers were combined to evaluate the benefit of carotid endarterectomy in randomized patients who had experienced symptoms related to arteriosclerotic stenosis of the carotid artery and who had received either best medical therapy alone or best medical therapy plus carotid endarterectomy. The outcome events were nonfatal and fatal stroke or death. A three-tier system identified and adjudicated the type, severity, and location of each stroke and the cause of any death. Data about patients submitted to carotid endarterectomy outside the trial were compiled at the Nonrandomized Data Center at the Mayo Clinic. Between December 27, 1987, and October 1, 1990, 1,212 patients were randomized, 596 to medical therapy, 616 to carotid endarterectomy. Cross-over from the medical to the surgical arm was low (4.2%). Patients eligible for the trial, but not randomized totaled 1,044.

The European Carotid Surgery Trial (ECST) was a multicentre trial of carotid endarterectomy for patients who, after a carotid territory non-disabling ischaemic stroke, transient ischaemic attack, or retinal infarct, were found to have a stenotic lesion in the relevant (ipsilateral) carotid artery. Over a 10 year period, 2518 patients had been randomised, and the mean follow-up was almost 3 years among the 2200 thus far available for analysis of the incidence of strokes that lasted more than 7 days.

The systematic review by Naylor et al included the review of primary results and secondary analyses from NASCET and ECST.

The study by Brooks et al, was designed to determine whether carotid angioplasty and stenting (CAS) is equivalent to carotid endarterectomy (CEA) in patients with symptomatic carotid stenosis >70% by a randomized, controlled trial in a community hospital and was a two-arm randomized study. This limitation of this study is that it was limited to a single institution, and a select "team" with experience in cerebral vascular disease and endovascular techniques, thus, cannot advocate that CAS replace CEA as a primary revascularization procedure in patients with symptomatic carotid stenosis. However, it is the first randomized prospective study to demonstrate that carotid angioplasty and stenting is equivalent to endarterectomy for the treatment of symptomatic carotid stenosis without added risk for major or minor stroke.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* There was consistency of results across studies included in the body of evidence. The studies evaluating different treatment methodologies used the NASCET or ECST methods for measuring the percentage of stenosis in patients, as a part of the inclusion criteria. The body of evidence shows that NASCET and ECST are solid methodological outcome studies, which have standardized the method of quantifying stenosis in several different types of imaging studies. The NASCET method for stenosis is to first assess the distal ICA beyond a notable stenosis degree to determine if it is partly decreased in diameter, and if so to recognize that a "near occlusion" situation exists for no ratio calculation, as it is fallacious, to declare as severe stenosis. If the distal ICA is not notably decreased, calculation of stenosis is done measuring the narrowest stenosis diameter from a view showing the worst carotid bulb stenosis, comparing it a measure of the normal ICA well beyond where the walls are parallel. The ECST method for stenosis is to measure the narrowest lumen, imagine where the unseen bulb walls might be, measure that diameter, and calculate a ratio of the two. These standardized criteria are still widely used by radiologists, in order to determine the severity of stenosis in symptomatic patients and also to determine whether or not patients will benefit from surgical intervention.

The proper measurement of carotid stenosis is essential to achieving the well established favorable results of carotid interventions such as endarterectomy and stenting. Each of the trials, included in the systematic reviews by Cina et al, Rothwell et al, Goldstein et al, Brooks et al, etc that have established the effectiveness of these procedures has determined a threshold degree of vascular narrowing above which the procedure is indicated. The NASCET method, based on the ratio of narrowest area of the vessel to the diameter of the distal lumen, is the most reliable and reproducible of these methods. In fact, the results of the other studies been recast in terms of the NASCET method, and this is the method used in subsequent studies. The carotid surgery and interventional procedures are only indicated when a properly measured stenosis exceeds the threshold, usually set at 70% in symptomatic patients. If a trial shows a positive outcome, clinical application of the trial data requires that the intervention population be as close as possible to the inclusion criteria of the succesful study. For this reason, measuring the carotid stenois on non invasive imaging studies in keeping with the NASCET methodology is beneficial. This is because the clinical benefit depended on the stenosis measurement. If the patients are rigorously selected, they are most likely to benefit from the decision to intervene, and less likely to incorrectly denied treatment that is likely to be beneficial.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net*

benefit - benefit over harms):

Patients with stenoses will benefit from physicians using a standardized method for stenosis calculation. Accuracy is extremely important as the calculation will justify the intervention selected for the patient, as evidence-based guidelines base treatment recommendations on the patient´s percentage of stenosis. No harms have been identified from using a standardized method of quantifying carotid stenosis.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? No

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: Not applicable

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: Not applicable

1c.13 Grade Assigned to the Body of Evidence: Not applicable

1c.14 Summary of Controversy/Contradictory Evidence: Several potential challenges to using the NASCET criteria include, failure to assess for near occlusion and measuring the still tapering bulb. Both of these have a wide range of non-compliance and both deal with creating a denominator for ratio calculation.

Fox AJ, Symons SP, Aviv RI, Howard P, Yeung R, Bartlett ES. Falsely claiming use of NASCET percentage stenosis method. Radiology. 2009 Nov;253:574-5.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

Cina CS, Clase CM, Haynes BR. Refining the indications for carotid endarterectomy in patients with symptomatic carotid stenosis: A systematic review. J Vasc Surg 1999;30:606-17.

Koelemay MJQ, Nederkoorn PJ, Reitsma JB, Majoie CB. Systematic Review of Computed Tomographic Angiography for Assessment of Carotid Artery Disease. Stroke 2004;35:2306-2312.

Rothwell PM, Gutnikov SA, Warlow CP. Reanalysis of the Final Results of the European Carotid Surgery Trial. Stroke 2003;34:514-523.

North American Symptomatic Carotid Endarterectomy Trial (NASCET) Steering Committee. North American Symptomatic Carotid Endarterectomy Trial: Methods, Patient Characteristics, and Progress. Stroke 1991;22:711-20.

European Carotid Surgery Trialists' Collaborative Group. MRC European Carotid Surgery Trial: interim results for symptomatic patients with severe (70–99%) or with mild (0–29%) carotid stenosis. Lancet. 1991; 337: 1235–1243.

Randomised trial of endarterectomy for recently symptomatic carotid stenosis: Final results of the mrc european carotid surgery trial (ECST). Lancet 1998;351:1379-1387.

Goldstein LB, Hasselblad V, Matchar DB, McCrory DC.Comparison and meta-analysis of randomized trials of endarterectomy for symptomatic carotid artery stenosis. Neurology. 1995 Nov;45(11):1965-70.

Naylor AR, Rothwell PM, Bell PR. Overview of the principal results and secondary analyses from the European and North American randomised trials of endarterectomy for symptomatic carotid stenosis. Eur J Vasc ndovasc Surg. 2003 Aug;26(2):115-29.

Brooks WH, McClure RR, Jones MR, Coleman TC, Breathitt L. Carotid angioplasty and stenting versus carotid endarterectomy: randomized trial in a community hosopital. J Am Coll Cardiol. 2001;38(6):1589-1595.

1c.16 Quote verbatim, the specific guideline recommendation *(Including guideline # and/or page #)*:

The panel recommended that the NASCET method of carotid stenosis measurement should be used when angiography is used to correlate the US findings.(1)

When MRA techniques are used for determining carotid stenosis, the report should reflect the methodology and reference the criteria for percent stenosis outlined in the NASCET. Also, the percent stenosis must be calculated using the distal cervical ICA diameter, where the walls are parallel, for the denominator. Similar to CTA, MRA with attention to the acquisition parameters and post-processing techniques can provide cross sectional measurements of stenosis that correlate with properly performed NASCET estimates of percent stenosis obtained with catheter angiography. In the setting of near occlusion, it may not be accurate to calculate percent stenosis ratios in the presence of post-stenotic arterial diamoeter decrease. Some MRA techniques may not be amenable to quantitative measurements, in which case qualitative assessment of stenosis should be provided.(2,3)

1c.17 Clinical Practice Guideline Citation: 1. Grant EG, Benson CB, Moneta GL, Alexandrov AV, et al. Carotid Artery Stenosis: Gray-Scale and Doppler US Diagnosis-Society of Radiologists in Ultrasound Consensus Conference. Radiology 2003;229:340-346.

2. American College of Radiology (ACR), American Society of Neuroradiology (ASNR), Society of NeuroInterventional Surgery (SNIS), Society for Pediatric Radiology (SPR). ACR-ASNR-SNIS-SPR practice guideline for the performance of pediatric and adult cervicocerebral magnetic resonance angiography (MRA). [online publication]. Reston (VA): American College of Radiology (ACR); 2010

3. American College of Radiology (ACR), American Society of Neuroradiology (ASNR). ACR-ASNR practice guideline for the performance and interpretation of cervicocerebral computed tomography angiography (CTA). [online publication]. Reston (VA): American College of Radiology (ACR); 2010.

1c.18 National Guideline Clearinghouse or other URL: 1. http://radiology.rsna.org/content/229/2/340.full.pdf 2. http://www.acr.org/SecondaryMainMenuCategories/quality_safety/guidelines/pediatric/cervicocerebral_mra.aspx 3. http://www.acr.org/SecondaryMainMenuCategories/quality_safety/guidelines/dx/head-neck/cervicocerebral_angio.aspx

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? No

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: Not applicable

1c.23 Grade Assigned to the Recommendation: Not graded

1c.24 Rationale for Using this Guideline Over Others: It is the PCPI policy to use guidelines, which are evidence-based, applicable to physicians and other health-care providers, and developed by a national specialty organization or government agency. In addition, the PCPI has now expanded what is acceptable as the evidence base for measures to include documented quality improvement (QI) initiatives or implementation projects that have demonstrated improvement in quality of care.

Based on the NQF descriptions for rating the evidence, what was the developer's assessment of the quantity, quality, and consistency of the body of evidence?

1c.25 Quantity: Moderate 1c.26 Quality: Moderate1c.27 Consistency: Moderate

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
0507_Evidence_MSF5.0_Data.doc,195_Performance_Score_Results.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
There is wide variation in the use of methods for stenosis calculation, which may also lead to variation in the appropriateness of carotid intervention. Since the degree of stenosis is an important element of the decision for carotid intervention, characterization of the degree of stenosis needs to be standardized. Requiring that stenosis calculation be based on a denominator of distal internal carotid diameter or, in the case of duplex ultrasound, velocity measurements that have been correlated to angiographic stenosis calculation based on distal internal carotid diameter, makes the measure applicable to both imaging and duplex studies.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
CMS Physician Quality Reporting System
This measure was included in the CMS Physician Quality Reporting System as measure #195 Stenosis Measurements in Carotid Imaging Studies from 2007 until now. There is a gap in care as shown by this data; 9.62% of patients reported on did not meet the measure.

Scores on this measure is N=48,218; Mean=28.93%
25th percentile: 66.67%
50th percentile: 94.12%
75th percentile: 100%

Exception Rate: This measure is not specified with exceptions. See attached performance data.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
There is sufficient performance data.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
After a search of the medical literature, no disparities have been identified in the area of stenosis measurement.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
The American College of Radiology advocates that performance measure data should, where possible, be stratified by race, ethnicity, and primary language to assess disparities and initiate subsequent quality improvement activities addressing identified disparities, consistent with recent national efforts to standardize the collection of race and ethnicity data.

A 2008 NQF report endorsed 45 practices including stratification by the aforementioned variables (1). A 2009 IOM report "recommends collection of the existing Office of Management and Budget (OMB) race and Hispanic ethnicity categories as well as

more fine-grained categories of ethnicity (referred to as granular ethnicity and based on one´s ancestry) and language need (a rating of spoken English language proficiency of less than very well and one´s preferred language of health-related encounters)." (2)

1. National Quality Forum Issue Brief (no. 10) Closing the Disparities Gap in Healthcare Quality with Performance Measurement and Public Reporting. Washington, DC: NWF, August 2008.

2. Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. March 2010. AHRQ Publication No. 10-0058-EF. Agency for Healthcare Research and Quality, Rockville, MD. Available at: http://www.ahrq.gov/research/iomracereport. Accessed May 25, 2010.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, Frequently performed procedure

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

The performance of carotid US and the interpretation of US results vary considerably from laboratory to laboratory. Interpretive criteria for carotid stenosis are either indiscriminately applied or the interpreters are uncertain about exactly how to make the diagnosis of carotid stenosis.(1)

More specifically, a study conducted by Byrd and colleagues evaluated routine practice in vascular laboratories across 26 countries participating in The Asymptomatic Carotid Surgery Trial (ACST) in order to determine the areas which are in need of future standardisation. With greater than 41 interpretation criteria reported and only 29% using a standardised Doppler angle, this study highlights the need for standardization in chartacterizing the degree of stenosis.(2)

Stroke is the second leading cause of death worldwide, after ischemic heart disease, and the most important cause of long-term disability. Eight-five percent of strokes are ischemic. Of these, two thirds are of carotid origin and one third are from the heart or other vessels. Fieschi et al found that, in conscious patients with acute ischemic strokes that necessitated admission to a stroke unit, 76% had angiographic evidence of complete occlusion of the internal carotid artery, the middle cerebral artery, or one of its branches. Most of these occlusions were thought to be embolic and of cerebrovascular origin.(3)

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. Grant EG, Benson CB, Moneta GL et al. Carotid artery stenosis: Gray-scale and doppler US diagnosis – Society of radiologists in ultrasound (SRU) consensus conference. Radiology. 2003; 229:340-346.

2. Byrd S, Robless P, Baxter A, Emson M, Halliday A. Carotid duplex ultrasonography: importance of standardisation. Int Angiol 1998; 17:248–254.

3. Cina CS, Clase CM, Haynes BR. Refining the indications for carotid endarterectomy in patients with symptomatic carotid stenosis: A systematic review. J Vasc Surg 1999;30:606-17.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Safety

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.acr.org/Quality-Safety/Quality-Measurement/Medicare-Value-Based-Programs/PQRS-Sample and
http://www.acr.org/Quality-Safety/Quality-Measurement/Performance-Measures

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Diagnostic_Imaging_Specifications-635884471388767886.docx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
There are no significant changes since last endorsement.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Final reports for carotid imaging studies that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
each final report for carotid imaging studies performed during a 12 consecutive month measurement period

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Numerator Definition:
Direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement - includes direct angiographic stenosis calculation based on the distal lumen as the denominator for stenosis measurement OR an equivalent validated method referenced to the above method (eg, for duplex ultrasound studies, velocity parameters that correlate with anatomic measurements that use the distal internal carotid lumen as the denominator for stenosis measurement)

Numerator Instructions:

This measure requires that the estimate of stenosis included in the report of the imaging study employ a method such as the North American Symptomatic Carotid Endarterectomy Trial (NASCET) method for calculating the degree of stenosis. The NASCET method calculates the degree of stenosis with reference to the lumen of the carotid artery distal to the stenosis. For duplex imaging studies the reference is indirect, since the degree of stenosis is inferred from velocity parameters and cross referenced to published or self-generated correlations among velocity parameters and results of angiography or other imaging studies which serve as the gold standard. In Doppler ultrasound, the degree of stenosis can be estimated using Doppler parameter of the peak systolic velocity (PSV) of the internal carotid artery (ICA), with concordance of the degree of narrowing of the ICA lumen. Additional Doppler parameters of ICA-to-common carotid artery (CCA) PSV ratio and ICA end-diastolic velocity (EDV) can be used when degree of stenosis is uncertain from ICA PSV. (Grant et al,
Society of Radiologists in Ultrasound, 2003)6.

A short note can be made in the final report, such as:
• "Severe left ICA stenosis of 70-80% by NASCET criteria" or
• "Severe left ICA stenosis of 70-80% by criteria similar to NASCET" or
• "70% stenosis derived by comparing the narrowest segment with the distal luminal diameter as related to the reported measure of arterial narrowing" or
• "Severe stenosis of 70-80% - validated velocity measurements with angiographic measurements, velocity criteria are extrapolated from diameter data as defined by the Society of Radiologists in Ultrasound Consensus Conference Radiology 2003; 229;340-346."

Documentation-Information populating the final report may reside in a dedicated field in the electronic health record (EHR) or picture archiving and communication system (PACS), however stenosis measurement information should be included in the final report in order to be readily accessible in all circumstances

FOR EHR SPECIFICATION:
No Current HQMF eCQM Available. We are in the process of developing full electronic measure specifications.

FOR ADMINISTRATIVE CLAIMS SPECIFICATIONS:
Report CPT II Code 3100F: Carotid imaging study report (includes direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement)

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
After a search of the medical literature, no disparities have been identified in the area of stenosis measurement.

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Children's Health, Populations at Risk, Senior Care

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
FOR EHR SPECIFICATION:
No Current HQMF eCQM Available. We are in the process of developing full electronic measure specifications.

FOR ADMINISTRATIVE CLAIMS SPECIFICATIONS:
Patient encounter during the reporting period (CPT): 36222, 70498, 70547, 70548, 70549, 93880, 93882

S.10. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
No Denominator Exclusions or Denominator Exceptions

S.11. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
None

S.12. Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*

We encourage the results of this measure to be stratified by race, ethnicity, gender, and primary language, and have included these variables as recommended data elements to be collected.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
To calculate performance rates:
1) Find the patients who meet the initial patient population (ie, the general group of patients that the performance measure is designed to address).
2) From the patients within the initial patient population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial patient population and denominator are identical.
3) From the patients within the denominator, find the patients who qualify for the Numerator (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator

If the patient does not meet the numerator, this case represents a quality failure.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
This measure is not based on a sample or survey.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
N/A

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Not applicable

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Individual

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility, Imaging Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
This is not a composite measure.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
0507_MeasureTesting_MSF5.0_Data.doc,nqf_testing_attachment_195_Final_Results_1.26.16.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0507
Measure Title: Stenosis measurement in carotid imaging studies
Date of Submission: 1/15/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or

whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
We used Medicare Part B administrative claims data for 2012-2014 for the reliability testing.

1.3. What are the dates of the data used in testing? 2012 -2014

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan

☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The numbers of physicians were 133,717 physicians
Among these physicians 128, 525 were claims and 5192 were from registry

- The data collection period was 2012-2014
- Data abstraction was performed in 2015

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

- Number of patients eligible were 6,877,159 (avg. per NPI is 51.43)
- Number of patients reported were 2,268,250 (avg. per NPI is 16.96)

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Data was only used for reliability testing.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

There are no SDS variables for this measure.

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

An assessment of measure reliability applying a reliability coefficient in the form of the signal to noise ratio (SNR). In SNR analysis, reliability is the measure of confidence in differentiating performance between physicians or other providers.[1] The

signal is the variability in measured performance that can be explained by real differences in physician performance and the noise is the total variability in measured performance. Reliability is then the ratio of the physician-to-physician variance to the sum of the physician-to-physician variance plus the error variance specific to a physician:

Reliability = Variance (physician-to-physician) / [Variance (physician-to-physician) + Variance (physician-specific-error]

A reliability equal to zero implies that all the variability in a measure is attributable to measurement error. A reliability equal to one implies that all the variability is attributable to real differences in physician performance. A reliability of 0.70 is generally considered a minimum threshold for reliability and 0.80 is considered very good reliability.

The SNR reliability testing is performed using a beta-binomial model. The beta-binomial model assumes the physician performance score is a binomial random variable conditional on the physician's true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates.

To estimate these parameters (Alpha and Beta) three steps were used:

1) Build a data file of the proper form for physician-to-physician variance estimation.

2) Use the Betabin SAS macro to estimate the physician-to-physician variance.[2]

3) Use the physician-to-physician variance estimate and the physician-specific information to calculate the physician specific reliability scores.

Reliability can be estimated at different points. The PCPI testing followed the convention of estimating reliability at two points: 1) at a minimum number of qualities reporting events per physician and 2) at the average number of quality reporting events per physician. We generally set the minimum number required as 10 events. Limiting the reliability analysis to only those physicians with a minimum number of events reduces the bias introduced by the inclusion of physicians without a significant numbers of events.

A physician level registry or claims database was used for extracting the relevant physician level information. Conditional on having measure data elements from a large and robust sample of physicians, a deidentified measure reliability analysis can be performed.

1. Adams JL, Mehrotra A, McGlynn EA, *Estimating Reliability and Misclassification in Physician Profiling*, Santa Monica, CA: RAND Corporation, 2010. www.rand.org/pubs/technical_reports/TR863. (Accessed on December 24, 2015.)
2. Wakelin I: MACRO Betabin. [http://www.sensory.org/library/files/SAS/betabin-v22.sas]

Data analysis included these fields:

Reporting Method	N (# of NPIs)	# Patients Eligible	Average Eligible Patients per NPI	# of Patients Reported	Average Patients reported per NPI	# Measure Met	% Measure Met (Mean)	# Exclusions	% Exclusions (Mean)

measure met > 0%								measure NOT met > 0%					
n	Min	p25	Median	p75	Max	# Measure NOT Met	% Measure NOT Met (Mean)	n	Min	p25	Median	p75	Max

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.g., *percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Physician to Physician variation

Label	*Estimate*	*StandardError*	*tValue*	*Probt*	*Alpha*	*Lower*	*Upper*
mu	0.2303	0.00114	202.11	<.0001	0.05	0.228	0.2325

alpha	0.08906	0.00059	150.56	<.0001	0.05	0.0879	0.0902
beta	0.2977	0.00203	146.97	<.0001	0.05	0.2938	0.3017

Summary of PQRS Reliability Score Stats Cumulative and by Year (2012 - 2014)

Year	*Number of Providers*	*Reliability p25*	*Reliability median*	*Reliability p75*	*Reliability mean*	*Reliability LCLM*	*Reliability UCLM*
2012	37142	0.81728	1	1	0.84524	0.84238	0.84809
2013	33493	0.63205	1	1	0.81781	0.81477	0.82085
2014	12953	0.99306	0.99722	0.99913	0.99473	0.99461	0.99484
All	83588	0.88581	1	1	0.85741	0.85561	0.85922

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The mean (CI), P25, median, P75 of the reliability score results are shown in the above table for all 3 years as well as by each year. Our mean (CI) reliability is 0.85741 (0.85561, 0.85922). A reliability of 0.80 is considered very good reliability. So according to the reliability testing analysis, the results demonstrated very good reliability.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*
See previous validity testing results

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
See previous validity testing results

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Since the previous submission, this measure underwent maintenance review by an expert panel. The review was completed in February 2015. New evidence was reviewed. The expert panel agreed that the measure remained valid based on existing and new evidence.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

2b4.4a. What were the statistical results of the analyses used to select risk factors?

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and*

compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; <u>if no empirical sensitivity analysis</u>, identify the approaches for handling missing data that were considered and pros and cons of each*)

Not applicable.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; <u>if no empirical analysis</u>, provide rationale for the selected approach for missing data*)

Not applicable.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment **Attachment:** Diagnostic_Imaging_Specifications.docx

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Clarifying definitions and instructions were added to the numerator based on feedback from the PQRS program.

This measure was found to be reliable and feasible for implementation.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
Not applicable.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement (Internal to the specific organization)	Payment Program Physician Quality Reporting System https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html?redirect=/pqri/ Value Modifier https://www.cms.gov/medicare/medicare-fee-for-service-payment/physicianfeedbackprogram/valuebasedpaymentmodifier.html Quality Improvement with Benchmarking (external benchmarking to multiple organizations) ACR NRDR Qualified Clinical Data Registry www.acr.org/qcdr

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

This measure has been used in the CMS Physician Quality Reporting Initiative/System 2007-2016.

PQRS#195 Radiology: Stenosis Measurement in Carotid Imaging Reports
This measure has one of the largest percentage point increases in clinical performance rate for PQRS 2010-2013.

- 2010 Performance Rate: 59.9%
- 2013 Performance Rate: 75.6%
- Percentage Change 2010-2013: 15.7%
- Number of eligible professionals reporting the measure in years 2010-2013: 6,716

These results include the claims, registry, and EHR reporting mechanisms. Results were restricted to a group of eligible professionals who reported the same measure from 2010 to 2013. This table includes measure performance regardless of whether eligible professionals who reported the measure met the satisfactory reporting requirement.

Reference: 2013 PQRS eRx Experience Report. April 8, 2015. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/AnalysisAndPayment.html

The ACR believes that the reporting of participation information is a beneficial first step on a trajectory toward the public reporting of performance results, which is appropriate since the measure has been tested and the reliability of the performance data has been validated. Continued NQF endorsement will facilitate our ongoing progress toward this public reporting objective. Additionally, the CMS Physician Compare website is phasing in quality measures over the next several years. Quality measures are tools that help measure health care processes and outcomes. These data are associated with the ability to provide high-quality health care and physician participation in quality programs such as PQRS and the Value Modifier.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
This is an accountability measure and used in the CMS quality and payment programs.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Trends in performance score for 2012-2014 shows an increase in the number of patients eligible, number of patients reported and met the measure as well as % met the measure. The mean of % measure met increased consistently from 16.85 (2012) to 24.91 (2013) and then 81.57 in 2014.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
The is significant improvement from 2012 to 2014 for this measure.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We are not aware of any unintended consequences related to this measurement.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
Yes

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American College of Radiology (ACR)
Co.2 **Point of Contact:** Judy, Burleson, jburleson@acr.org
Co.3 **Measure Developer if different from Measure Steward:** American College of Radiology (ACR)
Co.4 **Point of Contact:** Judy, Burleson, jburleson@acr.org, 464-4904-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
List of Work Group Members:
William Golden, MD (Co-Chair) (internal medicine)
David Seidenwurm (Co-chair) (diagnostic radiology)
Michael Bettmann, MD
Dorothy Bulas, MD (pediatric radiology)
Rubin I. Cohen, MD, FACP, FCCP, FCCM
Richard T. Griffey, MD, MPH (emergency medicine)
Eric J. Hohenwalter, MD (vascular interventional radiology)
Deborah Levine, MD, FACR (radiology/ultrasound)
Mark Morasch, MD (vascular surgery)

Paul Nagy, MD, PhD (radiology)
Mark R. Needham, MD, MBA (family medicine)
Hoang D. Nguyen (diagnostic radiology/payer representative)
Charles J. Prestigiacomo, MD, FACS (neurosurgery)
William G. Preston, MD, FAAN (neurology)
Robert Pyatt, Jr., MD (diagnostic radiology)
Robert Rosenberg, MD (diagnostic radiology)
David A. Rubin, MD (diagnostic radiology)
B Winfred (B.W.) Ruffner, MD, FACP (medical oncology)
Frank Rybicki, MD, PhD, FAHA (diagnostic radiology)
Cheryl A. Sadow, MD (radiology)
John Schneider, MD, PhD (internal medicine)
Gary Schultz, DC, DACR (chiropractic)
Paul R. Sierzenski, MD, RDMS (emergency medicine)
Michael Wasylik, MD (orthopedic surgery)

Diagnostic Imaging Measure Development Work Group Staff
American College of Radiology: Judy Burleson, MHSA; Alicia Blakey, MS
American Medical Association-convened Physician Consortium for Performance Improvement: Mark Antman, DDS, MBA; Kathleen Blake, MD, MPH; Kendra Hanley, MS; Toni Kaye, MPH; Marjorie Rallins, DPM; Kimberly Smuk, RHIA; Samantha Tierney, MPH; Stavros Tsipas, MA
National Committee for Quality Assurance: Mary Barton, MD

PCPI measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study must be equal contributors to the measure development process. In addition, the PCPI strives to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2007
Ad.3 **Month and Year of most recent revision:** 02, 2015
Ad.4 **What is your frequency for review/update of this measure?** These measures will be updated every 3 years.
Ad.5 **When is the next scheduled review/update for this measure?** 09, 2017

THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Limited proprietary coding is contained in the Measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. The AMA, ACR, the PCPI and its members disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT®) or other coding contained in the specifications.

Ad.7 **Disclaimers:** See copyright statement above.

Ad.8 **Additional Information/Comments:** Coding/Specifications updates occur annually. The ACR has a formal measurement review process that stipulates regular (usually on a three-year cycle, when feasible) review of the measures. The process can also be activated if there is a major change in scientific evidence, results from testing or other issues are noted that materially affect the integrity of the measure.

Provided by the developer to NQF via e-mail on Wednesday, March 16, 2016

We discovered a copy-and-paste error in the data included in the supplemental submission. Instead of including numbers for physicians with at least 10 patients (which is the dataset used for reliability analysis), we erroneously submitted the numbers on all physicians. The numbers in the table in the attachment sent Monday are correctly limited to physicians with at least 10 patients. We apologize for this error and the confusion resulting from it. We are attaching the Excel file with calculation results for all physicians and for physicians with at least 10 eligible patients.

Please see below for edits on our submitted language. The edits are to clarify the language and to correct the copy-and-paste error (copying all physicians table instead of physicians with less than 10 excluded). Also included is some supplemental material that hopefully clarifies and reinforces the submission.

1(b)

CMS Physician Quality Reporting System

This measure was included in the CMS Physician Quality Reporting System as measure #195 Stenosis Measurements in Carotid Imaging Studies from 2007 until now. There is a gap in care as shown by this data.

Scores on this measure for 2012-2014 (calculated using data from CMS):

N=~~48,218~~ 36863 physicians with at least 10 patients had a non-zero reporting rate. Across these physicians, 22.6% of physicians did not meet the measure (100% - 77.4% who met the measure). Across physicians with at least 10 patients and a performance rate greater than zero for the 3-year period 2012-2014, mean performance rate= 77.4%. (Performance rate column in attached Excel).

The performance rate quartiles for the same period 2012-2014 for physicians with at least 10 patients and performance rate >0 were as follows:

25th percentile: ~~66.67~~% 65.28%

50th percentile: ~~94.12~~% 90.91%

75th percentile: 100%

Supplemental information:

This measure has been included in the Physician Quality Reporting System since 2007 as Measure #195. Shown below are national average performance rates as reported in the CMS Report: 2013 Reporting Experience Including Trends (2007-2014) Physician Quality Reporting System and Electronic Prescribing (eRx) Incentive Program, APPENDIX, Table A27. Reporting and Performance Information by Individual Measure for the Physician Quality Reporting System (2010 to 2013).

Year	Average Performance Rate
2010	59.5%
2011	61.7%
2012	66.1%
2013	76.2%

The performance rate was calculated as the count of reported instances where performance was met (numerator) divided by the total number of reported instances that excluded reported exclusions (i.e., performance denominator).

(link: http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/PY2013-Prior-Year-Benchmarks-.pdf

While these rates do show a steady increase in performance, the 2013 score still indicates that 23.8% of patients reported on did not receive optimal care.

Performance rates calculated from the claims data we got from CMS show a similar picture:

Trends in performance score for 2012-2014 shows a decrease in the number of patients eligible; the number of patients reported slightly increased as did the measure performance rate. The performance rate was calculated as the count of reported instances where performance was met (numerator) divided by the total number of reported instances.

	# Patients Eligible	# of Patients Reported	Reporting rate	Average Patients reported per NPI	# Measure Met	Performance rate
2012	3,106,223	713,708	23.0%	19.22	490206	68.7%
2013	2,717,439	753,817	27.7%	22.51	595575	79.0%
2014	879,749	754,569	85.8%	58.25	634260	84.1%

Note: We erroneously reported an increase in number of patients eligible. The table is included here to show the data that were in the Word document sent Monday. Additionally, we have added the "Performance Rate" column. This corrects the performance scores reported from our previous statement : The mean of % measure met across all physicians (not just those with non-zero performance and at least 10 patients) increased consistently from 16.85 (2012) to 24.91 (2013) and then 81.57 (2014).

The 2014 rate indicates that 15.9% of patients reported on did not receive optimal care.

Additionally, the percentage of eligible professionals who could have reported the measure has remained low until 2014; reporting rate increased from 23.0% in 2012 to 85.8% in 2014. While increased reporting rate increases the potential for patients receiving optimal care, at the 2014 rates there were still approximately 125,000 patients who were potentially not receiving optimal care per the measure. Rates below are from PQRS participation data received from CMS for years 2012-2014.

Year Reporting Rate

2012 23.0%

2013 27.7%

2014 85.8%

From section 4b.

Trends in performance score for 2012-2014 shows a decrease in the number of patients eligible, increase in the number of patients reported and met the measure as well as % met the measure.

	# Patients Eligible	# of Patients Reported	Reporting rate	Average Patients reported per NPI	# Measure Met	Performance rate
2012	3,106,223	713,708	23.0%	19.22	490206	68.7%
2013	2,717,439	753,817	27.7%	22.51	595575	79.0%
2014	879,749	754,569	85.8%	58.25	634260	84.1%

The performance rate was calculated as the count of reported instances where performance was met (numerator) divided by the total number of reported instances. limiting to physicians with at least 10 patients. The performance rate increased consistently from 2012-2014:

68.7% (2012)

79.0% (2013)

84.1% (2014).

These figures are a correction of the statement below:

The mean of % measure met across all physicians (not just those with non-zero performance and at least 10 patients) increased consistently from 16.85 (2012) to 24.91 (2013) and then 81.57 (2014).

FROM SPREADSHEET ATTACHMENT

Measure 195

(No Exclusions)

	measure met > 0%					
	n	Min	p25	Median	p75	Max
All- Claims and Registry						
All 3 Years	48218	0.54	66.67	94.12	100	100
2012	14027	0.54	50	86.11	100	100
2013	16870	0.63	68.75	93.47	100	100
2014	17321	0.69	80.65	98.11	100	100
Claims						
All 3 Years	43117	0.54	66.67	94.29	100	100
2012	12664	0.54	50	85.31	100	100
2013	15267	0.63	69.09	93.75	100	100
2014	15186	0.69	81.69	98.7	100	100
Registry						
All 3 Years	5101	0.65	68.75	92.56	100	100
2012	1363	0.65	57.14	97.26	100	100
2013	1603	1.52	66.67	90	100	100
2014	2135	3.17	75.7	92.31	100	100

Measure 195

(Excluding providers if eligible patients were 10 or less)

	measure met > 0%					
	n	***Min***	***p25***	***Median***	***p75***	***Max***
All- Claims and Registry						
All 3 Years	36863	0.54	65.28	90.91	100	100
2012	11159	0.54	47.62	82.86	99.01	100
2013	12971	0.63	66.67	90	100	100
2014	12733	0.69	78.95	95.15	100	100
Claims						
All 3 Years	33059	0.54	65	90.91	100	100
2012	10032	0.54	47.17	81.82	98.17	100
2013	11737	0.63	66.67	90.38	100	100
2014	11290	0.69	79.79	96	100	100
Registry						
All 3 Years	3804	0.65	66.67	88.89	100	100
2012	1127	0.65	50.91	92.86	100	100
2013	1234	1.52	64.42	85.45	98.81	100
2014	1443	3.17	75	89.58	97.56	100

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 0661
Measure Title: Head CT or MRI Scan Results for Acute Ischemic Stroke or Hemorrhagic Stroke Patients who Received Head CT or MRI Scan Interpretation within 45 minutes of ED Arrival
Measure Steward: Centers for Medicare & Medicaid Services
Brief Description of Measure: This measure calculates the percentage of acute ischemic stroke or hemorrhagic stroke patients who arrive at the emergency department (ED) within two hours of the onset of symptoms and have a head computed tomography (CT) or magnetic resonance imaging (MRI) scan interpreted within 45 minutes of ED arrival. The measure is calculated using chart-abstracted data, on a rolling, quarterly basis and is publicly reported, in aggregate, for one calendar year. The measure has been publicly reported, annually, by CMS as a component of its Hospital Outpatient Quality Reporting (HOQR) Program since 2012.
Developer Rationale: Prompt brain imaging is a critical component of ED evaluation for patients with suspected acute stroke because it provides important information about the diagnosis, prognosis, and immediate and long-term treatment of these patients. A head CT or MRI scan is recommended to differentiate ischemic strokes, hemorrhagic strokes, and stroke mimics, and to identify appropriate candidates for tissue plasminogen activator (tPA), which is the gold standard for treating acute ischemic stroke (Jauch et al. 2013). Because the Food and Drug Administration (FDA) has approved tPA to be administered within three hours of symptom onset, expedited imaging can facilitate administration of the time-sensitive therapy for eligible patients (Cheng et al. 2015).

Numerator Statement: The number of acute ischemic stroke or hemorrhagic stroke patients who arrive at the ED within two hours of the onset of symptoms and have a head CT or MRI scan interpreted within 45 minutes of ED arrival.
Denominator Statement: The number of acute ischemic stroke or hemorrhagic stroke patients who arrive at the ED within two hours of the onset of symptoms and have a head CT or MRI scan ordered.
Denominator Exclusions: Studies are excluded for any patients under 18 years of age, patients who expired in the ED, or patients who left the ED against medical advice or discontinued care. Additionally, patients who do not arrive to the ED within two hours of symptom onset or who do not have a head CT or MRI scan ordered are excluded from the target population.

Measure Type: Process
Data Source: Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Jan 17, 2011 **Most Recent Endorsement Date:** Jan 17, 2011

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

Evidence for this process measure should demonstrate that, if stroke patients arrive at the ED within two hours of onset of symptoms and have a head imaging with a CT or MRI scan interpreted within 45 minutes of ED arrival, they will have better clinical outcomes without risk of hemorrhage.

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☒ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary and **Summary of prior review in 2010**

- The developer cited three articles that concluded that radiology report turnaround time can impact patient throughput times in the emergency department (Delfino, 2008); that decreasing radiology report turnaround times can have impacts across the facility and assist in reducing the length of stay and enhancing decision making capabilities for patient treatment plans (Marquez, 2005), and that efficiencies in throughput with tasks can lead to less diversion, less overcrowding, fewer elopements, and less financial loss (Falvo, 2007).
- The measure initially submitted for endorsement was a broad measure that assessed, for all patients, the median time from initial ordering of a head CT scan to reporting of the head CT scan results to ED staff. The Committee recommended limiting the denominator to stroke patients, expanding the definition of the measure to include MRI, and assessing percentage of patients who had a timely interpretation of a CT/MRI scan rather than assessing median time for the scan. The developer agreed to make these changes to the measure.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

The developer provided two guidelines from the American Heart Association/American Stroke Association for the early management of patients with acute ischemic stroke including endovascular treatment.

- Guideline #1 provides three recommendations for patients with acute cerebral ischemic symptoms that have not yet resolved.
    - Emergency imaging of the brain is recommended before initiating any specific therapy to treat acute ischemic stroke. In most instances, NECT [non-contrast-enhanced computed tomography] will provide the necessary information to make decisions about emergency management (p.18). (Unchanged from the previous guideline). **Class I: Level of Evidence A**
    - Either NECT or MRI is recommended before intravenous rtPA administration to exclude ICH [intracranial hemorrhage] (absolute contraindication) and to determine whether CT hypodensity or MRI hyperintensity of ischemia is present (p.18). (Revised from the 2009 imaging scientific statement). **Class I: Level of Evidence A**
    - In intravenous fibrinolysis candidates, the brain imaging study should be interpreted within 45 minutes of patient arrival in the emergency department by a physician with expertise in reading CT and MRI studies of the brain parenchyma (p. 3). (Revised from the previous guideline). **Class I: Level of Evidence C** (Consensus of opinion of the experts and/or small studies, retrospective studies, registries).
- Guideline #2 provides a focused update of the current recommendations for the endovascular treatment of acute ischemic stroke.
    - Emergency imaging of the brain is recommended before initiating any specific treatment for acute stroke. In most instances, nonenhanced CT will provide the necessary information to make decisions about

emergency management (p 29). (Unchanged from the 2013 guideline). **Class I, Level of Evidence A**

- The developer states that there is a broad consensus in the medical community supporting the recommendation that the brain imaging should be interpreted within 45 minutes of ED arrival and that "*the AHA Stroke Council asserts that a recommendation with Level of Evidence B or C does not imply that the recommendation is weak, as many important clinical questions addressed in the guidelines do not lend themselves to clinical trials*."
- The developer also provided six articles which they state support the measure's intent; however, the articles focus on the use and effectiveness of tPA in stroke patients. The focus of the measure is the interpretation of brain imaging within 45 minutes of arrival to the ED.

Exception to evidence – N/A

Guidance from the Evidence Algorithm

Process measure (Box 3)→Systematic Review of the evidence within the AHA/ASA guideline development (Box 4)→QQC partially available (Box 4)→ Class I recommendations (Evidence and/or general agreement that given treatment or procedure is beneficial, useful, and effective) implies that the evidence review concludes a moderate-high certainty that the net benefit is substantial (Box 5) → moderate (due to lack of detail on the quality and consistency of the evidence; Level C grade for 45-minute requirement)

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to timely provision and interpretation of CT/MRI imaging for stroke patients?*
- *Does empirical evidence support the requirement of interpretation of imaging studies within 45 minutes of ED arrival?*

Preliminary rating for evidence: ☐ ***High*** ☒ ***Moderate*** ☐ ***Low*** ☐ ***Insufficient***

Preliminary rating for evidence: ☒ ***Pass*** ☐ ***No Pass***

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance Measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer provided the following data for current performance:

	2012-2013	2013-2014
# facilities	918	959
# CT/MRIs performed (den)	16,817	17,108
Weighted mean performance	59.6%	67.4%
Standard deviation	99.9	91.3
Facility median	62.0%	71.0%
Range	0 - 100%	0-100%
10th percentile	24.0%	33.0%
25th percentile	42.0%	54.0%
90th percentile	88.0%	92.0%

Disparities:

- The developer provided the following information showing that race, sex, and facility characteristics played a role in determining whether a patient had a head CT or MRI interpreted within 45 minutes of ED arrival: African-American patients were less likely than White patients; Hispanic patients were less likely than non-Hispanic patients and female patients were less likely than male patients to have a head CT or MRI scan interpreted within 45 minutes of ED arrival.
- Those patients treated in facilities with fewer than 50 beds and those treated in major teaching facilities were less likely to have CT or MRI scan interpreted within 45 minutes of ED arrival.

Questions for the Committee:

- *Does the gap in care warrant a national performance measure?*
- *Do you think that there are opportunities for improvement for small or teaching hospitals? What about differences based on ethnicity, gender or race?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **Evidence presented is primarily about facility throughput. Rapid scan interpretation will rule out tPA in hemorrhagic stroke. There is community consensus.

**Expert consensus opinion arbitrarily cites 45 minutes from ED arrival to Interpreation time

1b. Performance Gap

Comments: **10th percentile - 33%, 90th percentile 92% Plenty of room for improvement.

Important disparities are discussed - no plan for elimination of them other than an expectation that the measure will "raise all boats."

**Some data on disparities presented.

In my opinion does not warrant national performance measure.

1c. High Priority (previously referred to as High Impact)

Comments: **Arbitrary time point is consensus opinion. Relatively easy to abstract.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Medical record abstraction (paper or electronic). This is not an eMeasure. Administrative claims are listed as a data source although this measure is manually abstracted.

- The numerator includes patients age 18 or older who were last known well within two hours of ED arrival and had a head CT or MRI ordered and interpreted within 45 minutes of ED arrival.
  - Numerator exceptions include: Patients with documentation of unable to determine (UTD) the following data elements: Data Last Known Well; Time Last Known Well; Arrival Time; Head CT Scan or MRI Interpretation Date; Head CT Scan or MRI Interpretation Time.
- The denominator includes patients age 18 or older who were last known well within two hours of ED arrival and had a head CT or MRI ordered.
  - The denominator excludes patients less than 18 years of age; patients who expired in the ED; patients

who left the ED against medical advice or discontinued care; patients who do not arrive to the ED within two hours of symptom onset or who do not have a head CT or MRI scan ordered.

- ICD-10 code and evaluation and management (E/M) codes included in Excel workbook and saved on Sharepoint.
- A detailed calculation algorithm is provided.
- Sampling is allowed; instructions provided.
- An electronic data collection tool is available from vendors or facilities can download the free CMS Abstraction & Reporting Tool (CART). Paper tools for manual abstraction are also available for the CART tool.
- The measure is specified at the hospital level of analysis.

Questions for the Committee

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- . Dataset used for testing included 92,633 cases from 2,985 hospital-associated outpatient services nationwide from January 1, 2012 to December 31, 2012. Military or Veterans Affairs and Critical Access Hospitals (CAH) were excluded.
- Data element validity testing was performed and counted for data element reliability as well – see validity testing section.

Describe any updates to testing:

- Reliability testing of the measure score conducted – see below.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing:

- The dataset used for testing included cases submitted from 958 facilities to Hospital Compare from January 1, 2014 – December 31, 2014. The sample included 17,162 denominator cases (initial population) and 11,741 numerator cases (CT/MRI interpretation w/in 45 minutes of ED arrival).
- The developer calculated the signal-to-noise ratio using a beta-binomial model for each facility meeting the minimum case count (10). *[Note: Ten is the minimum number of cases required for public reporting. It is unclear whether the measure itself is limited to facilities with 10 or more cases; if it is not, then testing was not conducted with the measure as specified.]* A *signal-to-noise approach is an appropriate methodology.*

Results of reliability testing:

- Reliability scores ranged from **0.62** to **1.00**. The median reliability score was **0.77**. A value of 0.7 is often regarded as a minimum acceptable reliability value.

Questions for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*
-

Guidance from the Reliability Algorithm:

Precise specifications (Box 1)→ empirical testing as specified (Box 2)→ reliability testing conducted with computed performance measure score (Box 4) → signal-to-noise analysis conducted (Box 5)→Moderate certainty or confidence that the performance measure scores are reliable (Box 6a)→ recommend Moderate

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Do you believe the specifications—including the 45-minute requirement—are consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

- Face validity of the measure score was systematically assessed through surveying an expert working group. The developer stated that "All members agreed or strongly agreed that patients who have a head CT scan or MRI ordered and interpreted within 45 minutes of ED arrival can be accurately captured using chart-abstracted data. Similarly, they agreed or strongly agreed that NQF #0661 successfully assesses the timely interpretation of head imaging for acute ischemic and hemorrhagic stroke patients. The respondents generally support the face validity of NQF #0661." NQF guidance requires that face validity of the measure score must explicitly address whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. It does not appear that the face validity assessment conducted by the developer conforms to NQF's requirements.
- Data element validity was assessed calculating the raw agreement rate between data from two different abstractors for 12 critical data elements. The developer stated that "*Overall, the agreement rates were high. The agreement rates for all data elements except three were higher than 90%. These three data elements were "Last Known Well" (79.8%), "Last Known Well Time" (74.6%), and "Scan Interpretation Time" (65.2%). The kappa statistic for the two dichotomous ("Last Known Well" and "Scan Ordered") data elements reflected moderate to almost perfect agreement even after kappa adjusts for agreement by chance alone*."

Describe any updates to validity testing: Additional data element testing is presented.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**
☒ **Empirical validity testing of the measure score**

Validity testing method:

- Empirical validity of the measure performance score of eight critical data elements was assessed by calculating a kappa statistic assessing the agreement between facility abstraction and auditor (CDAC) abstraction from 774

cases from April 1, 2014-March 31, 2015.

Validity testing results:

- The kappa statistic for numerator cases equals 0.52, with a p-value of less than 0.001. The kappa statistic for denominator cases equals 1.00, with a p-value of less than 0.001.
- The developer does not provide the validity testing results for the eight data elements but instead states, "The rate of agreement, by data element, ranged from 52.7% to 98.4%. The rate of agreement was strong for dichotomous variables, as well as those based on administrative data. This included variables, such as Scan Ordered, Discharge Code, and Scan Order Date. Agreement was moderate for clinical variables related to time, such as Time Last Known Well and Scan Order Time."
- The developer states that results of the face-validity assessment indicate that a diverse group of stakeholders support the validity of the measure. NQF guidance requires that face validity of the measure score must explicitly address whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. It does not appear that the face validity assessment conducted by the developer conforms to NQF's requirements.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*

2b3-2b7. Threats to Validity

2b3. Exclusions

Overall Occurrence and Distribution across Facilities for Measure Exclusions and Exceptions

Data Element	Denominator Exclusion or Numerator Exception?		Overall Occurrence		Distribution Across Facilities (%)		
	Denominator Exclusion	Numerator Exception	N	%	25th	50th	75th
Discharge Code	X		4,627	4.4	0.0	2.5	6.7
Head CT Scan or MRI Order	X		3,537	3.3	0.0	1.6	5.3
Last Known Well	X		44,328	41.9	28.6	41.4	52.6
Date Last Known Well		X	100	0.1	0.0	0.0	0.0
Time Last Known Well		X	1,506	1.4	0.0	0.0	0.0
Arrival Time		X	5	0.0	0.0	0.0	0.0
Last Known Well Minutes	X		25,170	23.8	14.3	22.2	30.4
Head CT Scan or MRI Interpretation Date		X	87	0.1	0.0	0.0	0.0
Head CT Scan or MRI Interpretation Time		X	373	0.4	0.0	0.0	0.0
Total Denominator Exclusions	4 exclusions	-	77,662	73.40	-	-	-
Total Numerator Exceptions	-	5 exceptions	2,071	2.00	-	-	-
Total Removed from the Denominator or Numerator	9 exceptions and exclusions		79,733	75.40	-	-	-

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*
- *Do you think that the high level of exclusions from the denominator is of concern?*
- *With the high percentage of exceptions and exclusions do you think this measure is can distinguish good quality from poor quality?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Using data from 2014, the developer tested the statistical significance of the difference between facility performance scores and the mean performance value for 958 facilities meeting public-reporting requirements.
- Results of the analysis indicated that the performance of 4.5% of the 958 facilities (n=43) was statistically significantly different from the average performance rates across the 958 facilities.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

According to the developer this measure only uses only one set of specifications.

-

2b7. Missing Data

- Missing data is not reported or adjusted for this measure but the developer points out that if data are missing, the case will be rejected. While abstractors cannot submit missing data, they submit a value of unable to determine (UTD) for select data elements. Depending on the data element the case is then either excluded from the denominator or excepted from the numerator.
- Exclusions criteria and percent excluded can be found in the table including the rate for UTD. All UTD were excluded from the numerator but not the denominator. The percentages ranged from 0.1- 1.4%.

Guidance from the Validity Algorithm: Specifications consistent with evidence (Box 1) → potential threats to validity assessed (Box2) →empirical validity testing (Box 3) → no empirical testing at the measure score level (Box 6) → validity testing conducted with patient-level data elements (Box 10) →percent agreement of the critical data elements assessed (Box 11)→Moderate certainty or confidence that the data used in the measure are valid (Box 12a)→Moderate
(Note: Moderate is the highest rating possible with data element validity only)

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **Face Validity only - no inclusion of improved outcome data.

**See above.

**Data elements are all clear.

2a2. Reliability Testing

Comments: **Face. The other validities mentioned are really reliabilities. There is no predictive validity!

**relatively easy metric to assess. Amenable to modifying process.

**.62-1.00 - median .77

2b2. Validity Testing

Comments: **There are several exclusions and exemptions identified - all seem appropriate.

Meaningful data - 45% of facilities

Missing data are just rejected cases

**Yes.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **.62-1.00 - median .77

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Administrative claims, Electronic Clinical Data, EHRs, Paper: An electronic data collection tool is made available from vendors or facilities or from CMS Abstraction & Reporting Tool (CART). Some data elements are in defined fields in electronic sources.
- Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Data are in the record.

Data extraction burden?

**Time stamp of arrival not always accurate in electronic record.

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- This measure is publically reported through the CMS HOQR Program, a pay for quality data reporting program implemented by CMS for outpatient hospital services. Hospital quality of care information gathered through the HOQR Program is publicly available on the Hospital Compare website.

Improvement results

- The developer reports that the median rate of head CT or MRI scans for acute ischemic or hemorrhagic stroke patients that are interpreted within 45 minutes of ED arrival, who arrived at the ED within two hours of the known onset, has increased 14.5% from 62.0% in 2012 to 71.0% in 2014.

Unexpected findings (positive or negative) during implementation

- The developer states that has been wide variation in facility performance from the inception of public reporting through September 2014. While median performance is improving, there is an ongoing opportunity for improvement in facility performance. Measure score validity is moderate with some differences in how facilities and Clinical Data Abstraction Center (CDAC) identify the numerator, suggesting that clearer abstraction guidance could improve the validity of the publically reported values.

- The developer states that many facilities are not meeting the minimum case count requirements for public reporting (more than ten cases with complete records after the application of measure exclusion criteria). Because the eligible denominator population is large (~540,000 ED cases with a primary diagnosis of stroke [HCUP 2012]), this is likely caused by 1) the small sampling requirements, and 2) application of exclusion criteria, which remove many eligible cases from calculation.

Potential harms: There was no potential harm reported.

Feedback :

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*
- *Are there opportunities to increase the number of reported counted on cases captured for this measure?*

- **Preliminary rating for usability and use: ☒ High ☐ Moderate ☐ Low ☐ Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Data are available in public and accountability forums. Results have improved since initial endorsement!
**Unintended consequence-rush to meet metrics may loose specificity of initial dianosis.

Criterion 5: Related and Competing Measures

Measure NQF #0661

Related Measures:

- The measure NQF #0437 (used in the Hospital Inpatient Quality Reporting [HIQR] Program) is similar to NQF #0661 (HOQR), the two measures serve different target populations and purposes: the HOQR measure focuses on imaging in the ED setting, while the HIQR measure focuses on administration of thrombolytic therapy in an inpatient setting.
- Some key data elements (i.e., Last Known Well, Date Last Known Well, Time Last Known Well, and Arrival Time) are shared by both of these measures.
- Developer stated that the measure maintenance team for the HOQR measure (NQF #0661) incorporates specification updates added by the measure maintenance team for the HIQR measure (NQF #0437) to maintain harmonization (e.g., updates to the appropriate ICD-10 codes to determine measure inclusion).

Harmonization:
Measures are not harmonized

Pre-meeting public and member comments

No comments were submitted for this measure during the pre-meeting comment period.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0661

Measure Title: Head CT or MRI Scan Results for Acute Ischemic Stroke or Hemorrhagic Stroke Patients who Received Head CT or MRI Scan Interpretation Within 45 Minutes of ED Arrival

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 1/15/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep

process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☐ Health outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Head CT or MRI Scan Results for Acute Ischemic Stroke or Hemorrhagic Stroke Patients who Received Head CT or MRI Scan Interpretation Within 45 Minutes of ED Arrival

☐ Structure: Click here to name the structure

☐ Other:

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

This measure is not a health outcome/PRO performance measure.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

This measure is not a health outcome/PRO performance measure.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

According to the Centers for Disease Control, stroke is the fifth leading cause for death (Kochanek et al. 2014). Prompt brain imaging is a critical component of an acute stroke patient's ED evaluation because it provides

important information about the diagnosis, prognosis, and immediate and long-term treatment of potential stroke patients. In particular, computed tomography (CT) or magnetic resonance imaging (MRI) can identify contraindications for time-sensitive treatment such as fibrinolysis. Once the appropriate therapy is determined, guidelines recommend that treatment be initiated without delay because the likelihood of favorable outcome is directly linked to the time-to-treatment (See Guideline #1 in Section **1a.4.a**) (Jauch et al., 2013). For example, fibrinolysis with intravenous recombinant tissue plasminogen activator (tPA or rtPA), the gold standard for acute ischemic stroke, has been approved by the Food and Drug Administration (FDA) to be administered within three hours of symptom onset (Cheng and Kim 2015). Although it has been shown to improve functional outcomes at three to six months when given within three hours of ischemic stroke onset for patients who meet eligibility criteria, there is evidence that a shorter time-to-treatment is associated with reduced mortality and symptomatic intracranial hemorrhage, and higher rates of independent ambulation at discharge and discharge to home (Saver et al. 2013). Accordingly, #0661 requires that a head CT scan or MRI be interpreted within 45 minutes for patients who are within two hours of symptom onset in order to ensure that eligible tPA candidates receive the time-sensitive treatment within the recommended three hour window.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Two clinical practice guidelines are provided based on their relevance to the measure. The first guideline, released in 2013 by the American Heart Association (AHA) and the American Stroke Association (ASA), evaluates the early management of patients with acute ischemic stroke. The second AHA/ASA guideline, released in 2015, is a focused update of the 2013 guideline with an emphasis on endovascular treatment. Citations for the two guidelines follow:

1 –.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Guideline 1 provides recommendations for patients with acute cerebral ischemic symptoms that have not yet resolved. Three recommendations support the measure's clinical intent.

Guideline 1 Recommendations: American Heart Association/ American Stroke Association

A. Emergency imaging of the brain is recommended before initiating any specific therapy to treat acute ischemic stroke. In most instances, NECT [non-contrast-enhanced computed tomography] will provide the necessary information to make decisions about emergency management. (Unchanged from the previous guideline). (Class I, Level of Evidence A; pg. 18).
B. Either NECT or MRI is recommended before intravenous rtPA administration to exclude ICH [intracranial hemorrhage] (absolute contraindication) and to determine whether CT hypodensity or MRI hyperintensity of ischemia is present. (Revised from the 2009 imaging scientific statement). (Class I, Level of Evidence A; pg. 18).
C. In intravenous fibrinolysis candidates, the brain imaging study should be interpreted within 45 minutes of patient arrival in the emergency department by a physician with expertise in reading CT and MRI studies of the brain parenchyma. (Revised from the previous guideline) (Class I, Level of Evidence C; pg. 3).

Guideline 2 provides a focused update of the current recommendations for the endovascular treatment of acute ischemic stroke. One recommendation supports the measure's clinical intent.

Guideline #2 Recommendation: American Heart Association/ American Stroke Association

A. Emergency imaging of the brain is recommended before initiating any specific treatment for acute stroke. In most instances, nonenhanced CT will provide the necessary information to make decisions about emergency management. (Unchanged from the 2013 guideline). (Class I, Level of Evidence A; pg. 29).

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

All relevant recommendations from *Guideline 1* received a Class I designation. The evidence (level of evidence A) strongly and unambiguously support the recommendations to perform either an NECT or MRI before initiating treatment (and specifically before administering tPA) for acute ischemic stroke. Additionally there is a broad consensus in the medical community (level of evidence C) supporting the recommendation that the brain imaging should be interpreted within 45 minutes of ED arrival. The AHA Stroke Council asserts that a recommendation with Level of Evidence B or C does not imply that the recommendation is weak, as many important clinical questions addressed in the guidelines do not lend themselves to clinical trials. Despite a limited pool of randomized control trial, there may be clear clinical consensus that a particular test or therapy is useful or effective. Collectively, the evidence supports these recommendations, demonstrating consensus within the clinical community that patients eligible for fibrinolysis should receive emergency imaging of the brain that should be interpreted within 45 minutes of ED arrival by a qualified physician.

The Class I recommendation from *Guideline 2* relates to the importance of expedited imaging, indicating consensus within the clinical community that emergency imaging is recommended regardless of the specific treatment being considered for acute stroke.

The following grading scale applies to recommendations from *Guideline 1*:

Recommendation A: Class I: Usefulness/efficacy is well established by evidence/opinion.

Recommendation B: Class I: Usefulness/efficacy is well established by evidence/opinion.

Recommendation C: Class I: Usefulness/efficacy is well established by evidence/opinion.

The following evidence scales apply to recommendations from *Guideline 1*:

One class of recommendations: Class I

Class I: Evidence and/or general agreement that given treatment or procedure is beneficial, useful, and effective.

Two levels of evidence: Level A and Level C.

Level A: Data derived from multiple randomized clinical trials or meta-analyses.

Level C: Consensus of opinion of the experts and/or small studies, retrospective studies, registries.

The following grading scale applies to recommendations from *Guideline 2*:

Recommendation A: Class I, Level A: Usefulness/efficacy is well established by evidence/opinion.

The following evidence scales apply to recommendations from Guideline 2:

Two classes of recommendations: Class I and Class IIb

Class I: Benefit >>> Risk. Procedure/treatment SHOULD be performed/administered.

One level of evidence: Level A.

Level A: Data derived from multiple randomized clinical trials or meta-analyses.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

All relevant information about the grades and associated definitions for the two guidelines provided has been included in Section **1a.4.3.**

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Citations and URLs are the same as those noted in Section **1a.4.1**.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

This measure is not based on a United States Preventive Services Task Force Recommendation.

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

This measure is not based on a United States Preventive Services Task Force Recommendation.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

This measure is not based on a United States Preventive Services Task Force Recommendation.

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

This measure is not based on a United States Preventive Services Task Force Recommendation.

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

This measure is not based on a United States Preventive Services Task Force Recommendation.

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Guidelines are evidenced based; details are provided in Section **1a.7**.

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Guidelines are evidenced based; details are provided in Section **1a.7**.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

Methodologic Approach for the Systematic Review that Supports *Guideline #1*:

Members of the writing committee were appointed by the ASA's Stroke Council's Scientific Statement Oversight Committee, representing various areas of medical expertise. Strict adherence to the AHA conflict of interest policy was maintained throughout the consensus process. Panel members were assigned topics relevant to their areas of expertise, reviewed the stroke literature with emphasis on publications since the prior guidelines in 2007, and drafted recommendations in accordance with the AHA Stroke Council's Level of Evidence grading algorithm. To develop consensus, the writing committee was polled on the wording, classes of recommendations, and levels of evidence assigned to each recommendation. The standard AHA algorithm for classifying recommendations and levels of evidence focuses on therapeutic questions and, consequently, emphasizes evidence from randomized clinical trials.

Methodologic Approach for the Systematic Review that Supports *Guideline #2*:

Members of the writing committee were appointed by the ASA Stroke Council's Scientific Statement Oversight Committee, representing various areas of medical expertise. Panel members were assigned topics relevant to their areas of expertise, reviewed the stroke literature with emphasis on publications since the prior guidelines, and drafted recommendations in accordance with the AHA's Stroke Council's Level of Evidence grading algorithm. Due to the wide scope of the guidelines, the panelists were assigned as a primary or secondary author for the individual sections. In some cases in which clinical trial research was unavailable, the panelists provided recommendations based on pathophysiological reasoning and expert practice experience.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

Grade for the evidence provided from *Guideline #1* can be found in Section **1a.4.3.**

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Grade for the evidence provided from *Guideline #1* can be found in Section **1a.4.3.**

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range:

Guideline #1: It is inferred that the time period covered by the body evidence is 2007 – 2012 since Recommendation C (brain imaging interpreted within 45 minutes of ED arrival) has been updated since the previous guideline published in 2007 and *Guideline 1* was published in January 2013.

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

-----look at evidence Guideline #1 does not indicate the specific number or type of study designs included in the body of evidence; however, two of the recommendations are Level A, which are based on data from multiple randomized clinical trials or meta analyses, and the third recommendation is Level C, which is based on consensus opinion of experts, case studies, or standard of care. The imaging recommendations referenced 228 unique citations with evidence from 3 systematic reviews, 5 guidelines, 20 randomized control trials, and 173 observational studies.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Guideline #1 provides three Class I recommendations, indicating that the benefits clearly outweigh the risks and the recommendation can be applied to most patients in most circumstances. The two Level A recommendations are based on randomized control trials (RCTs) with no important limitations or exceptionally strong evidence from observational studies, and further evidence is unlikely to change the confidence in the estimate of the the effect. A Level A study of diagnostic or prognostic accuracy would be a prospective cohort survey. Investigators would start with a group of patients suspected of having a disease (the cohort). The diagnostic test would be performed on this cohort. Some patients would have a positive test, others a negative test. The cohort would then have the actual presence or absence of the disease determined by an independent reference standard (the gold standard). Quantitative measures of the diagnostic accuracy of the test (or predictor) such as the sensitivity or specificity could then be calculated. For a study to be graded Level A, an investigator who is unaware of the results of the diagnostic test (presence or absence of the prognostic predictor) should apply the reference standard to determine the true presence of the disease (outcome). The third recommendation (Level C) is based on observational studies, case series, or indirect evidence, such as the consensus opinion of experts or standard of care.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

Guideline #1 does not provide details about the estimates of benefit and consistency across studies in the body of evidence.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Guideline #1 does not provide details about potential harms associated with expedited brain imaging of acut stroke patients that were identified in the body of evidence.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

In addition to the two guidelines cited above, a review of the clinical literature was conducted during the measure contractor's annual review of the literature for additional evidence and/or new studies that relate to the measure. Citations and summaries for the six items included in this review can be found in Section **1a.8.2**. Some of these six studies have been published since the period of guideline development. Results cited in these studies are consistent across studies and with the guidelines cited above.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

In addition to the two guidelines cited above, a review of the clinical literature and related policy was conducted during the measure contractor's annual review of the literature for additional evidence and/or new studies that support the measure's intent. The measure contractor identified relevant peer-reviewed publications by searching the PubMed MEDLINE database from January 1, 2013 to February 15, 2015, limiting included results to those published in the English language and that had abstracts available in PubMed. The search initially identified 508 articles; a further review by the contractor's clinical and measure-development team resulted in the inclusion of six articles in the body of evidence below. Citations and summaries for the six items included in this review can be found in Section **1a.8.2**.

1a.8.2. Provide the citation and summary for each piece of evidence.

Edlow J, Smith E, Stead L, Gronseth G, et al. American College of Emergency Physicians, American Academy of Neurology. Clinical policy: use of intravenous tPA for the management of acute ischemic stroke in the emergency department. Ann Emerg Med. 2013; 61(2):225-43.
A joint writing panel of the American College of Emergency Physicians and the American Academy of Neurology reviewed literature, graded evidence, and made recommendations based on the strength of available data. Recommendations were developed to help clinicians answer questions including: (1) Is intravenous tissue plasminogen activator (tPA) safe and effective for acute ischemic stroke patients if given within 3 hours of symptom onset? (2) Is intravenous tPA safe and effective for acute ischemic stroke patients treated between 3 to 4.5 hours after symptom onset? The authors indicate that although the time window for tPA may have been

lengthened to 4.5 hours, patient outcomes are optimized by the earliest possible intervention after brain imaging and clinical evaluation.

Schwamm L, Ali S, Reeves M, Smith E, et al. Temporal trends in patient characteristics and treatment with intravenous thrombolysis among acute ischemic stroke patients at Get With The Guidelines-Stroke hospitals. Circulation. Cardiovascular Quality and Outcomes. 2013; 6(5):543-9.

Schwamm et al. analyzed all acute ischemic stroke patients arriving within two hours of symptom onset and treated with tPA within three hours of symptom onset from 2003 to 2011 in the American Heart Association's Get with the Guideline-Stroke (GWTG-Stroke). A univariate analysis revealed that tPA use increased over time, particularly in those aged older than 85 years, nonwhite, and with milder strokes. Additionally, door-to-image and door-to-tPA times also improved. Schwamm et al. found that the frequency of intravenous tPA use among all acute ischemic stroke patients nearly doubled from 2003 to 2011.

Lang C, Bland M, Cheng N, Corbetta M, et al. A case-control study of the effectiveness of tissue plasminogen activator on 6 month patients--reported outcomes and health care utilization. Journal of Stroke and Cerebrovascular Diseases: The Official Journal of National Stroke Association. 2014; 23(10):2914-9.

Lang et al. performed a cohort study to examine the benefit of tPA on patient-reported outcomes and health care utilization on 6-month stroke patients by analyzing patients who received tPA as part of usual stroke management and patients who would have received tPA had they arrived to the hospital within the therapeutic time window. Data were collected from surveys 6 months after stroke using standardized patient-reported outcome measures and questions about health care utilization. Demographic and medical data were acquired from hospital records. The tPA (n = 78) and control (n = 156) groups were matched across variables, except for stroke severity, which was better in the control group; subsequent analyses controlled for this mismatch. Patients who received tPa were compared with those who would have received tPA had they arrived to the hospital within the therapeutic window. The tPA group reported better physical function, communication, cognitive ability, depressive symptomatology, and quality of life/participation compared with the control group and fewer people in the tPA group reported skilled nursing facility stays, emergency department visits, and rehospitalizations after their stroke. Lang et al. found that the use of tPA provides a large benefit to the daily lives of people with ischemic stroke.

Choi J, Jang M, Kang K, et al. Comparative effectiveness of standard care with IV thrombolysis versus without IV thrombolysis for mild ischemic stroke. Journal of the American Heart Association. 2015; 4(1):e001306.

Choi et al. conducted an observational registry-based study to evaluate the comparative effectiveness of standard care with intravenous thrombolysis (IVT) versus without IVT in mild stroke patients. Choi et al. identified patients with acute ischemic stroke who presented within 4.5 hours of symptom onset and had National Institutes of Health Stroke Scale Scores of 5 or higher. Of 13,117 patients with stroke who were hospitalized between April 2008 and May 2012, 1,386 met eligibility criteria and 194 were treated with IVT. Choi et al. found that standard care with IVT is more effective than not receiving IVT in mild ischemic stroke patients, and there is a statistically insignificant risk of symptomatic hemorrhagic transformation.

Ciccone A, Valvassori L, Nichelatti M, Sgoifo A, et al. Endovascular treatment for acute ischemic stroke. The New England Journal of Medicine. 2013; 368(10):904-13.

Ciccone et al. conducted a randomized control trial of 362 patients with acute ischemic stroke who arrived within 4.5 hours after onset to compare the clinical efficacy of endovascular therapy and intravenous tPA. The median time from stroke onset to start of treatment was 3.75 hours for endovascular therapy and 2.75 hours for intravenous tPA ($p = 0.001$). There were no significant differences between groups in the rates of other serious adverse events or the case fatality rate, suggesting that endovascular therapy is not superior to the current gold standard of tPA.

Haršány M, Tsivgoulis G, Alexandrov A. Intravenous thrombolysis in acute ischemic stroke: standard and potential future applications. Expert Review of Neurotherapeutics. 2014; 14(8):879-92.

Haršány et al. reviewed studies providing evidence that intravenous thrombolysis with tPA improves early functional outcomes in acute ischemic stroke patients. Additionally, successful use of intravenous thrombolysis is dependent upon the organization of the treatment team and it should be standard that intravenous tPA be administered within 4.5 hours of the onset of stroke symptoms.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

NQF_0661_Measure_Evidence_Form-635882753624388714.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*

Prompt brain imaging is a critical component of ED evaluation for patients with suspected acute stroke because it provides important information about the diagnosis, prognosis, and immediate and long-term treatment of these patients. A head CT or MRI scan is recommended to differentiate ischemic strokes, hemorrhagic strokes, and stroke mimics, and to identify appropriate candidates for tissue plasminogen activator (tPA), which is the gold standard for treating acute ischemic stroke (Jauch et al. 2013). Because the Food and Drug Administration (FDA) has approved tPA to be administered within three hours of symptom onset, expedited imaging can facilitate administration of the time-sensitive therapy for eligible patients (Cheng et al. 2015).

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Analysis of facility-level data from Hospital Compare downloadable files indicates that there is variation in the use of head CT and MRI scans within 45 minutes of ED arrival for patients with a principal diagnosis of acute ischemic or hemorrhagic stroke. During the October 2012 through September 2013 data collection period, performance scores ranged from 0.0% to 100.0%, with a weighted mean of 59.6%. During the October 2013 through September 2014 data collection period, performance scores ranged from 0.0% to 100.0%, with a weighted mean of 67.4%, representing a 13.1% increase in the weighted mean of performance scores between the two time periods.

The data presented below represent performance scores for the 918 and 959 facilities whose denominator counts met minimum case count requirements during the October 2012 through September 2013 and October 2013 through September 2014 data collection periods, respectively. Publicly available data, at the facility level, was first included on Hospital Compare downloadable files for the data collection period of October 2012 through September 2013. To conduct longitudinal analysis, the most recent and complete twenty-four months of data were used: October 2012 through September 2013 and October 2013 through September 2014. These data collection periods contain the most recent data that are available from two distinct time periods.

Further details on the descriptive statistics for longitudinal facility performance are included below:

	Data Collection Period		Percentage Point Change (1)	
	Oct.2012-Sept. 2013		Oct. 2013-Sept. 2014	Oct. 2012-Sept. 2014
Facilities	918	959	-	
Minimum Value	0.0	0.0	-	
1st Percentile	4.0	8.0	4.0	
5th Percentile	15.0	23.0	8.0	
10th Percentile	24.0	33.0	9.0	
25th Percentile	42.0	54.0	12.0	
Median	62.0	71.0	9.0	
75th Percentile	78.0	83.0	5.0	
90th Percentile	88.0	92.0	4.0	
95th Percentile	92.0	96.0	4.0	
99th Percentile	100.0	100.0	0.0	
Maximum Value	100.0	100.0	0.0	

Weighted Mean Performance (Standard Deviation) 59.6 (99.9) 67.4 (91.3) 7.8

Number of CT and MRI scans performed (Denominator) 16,817 17,108 -
(1) Note that this value represents the percentage point change, not the percentage increase/decrease.

From the inception of public reporting through September 2014 data collection, there has been wide variation in facility performance. During the October 2012 through September 2013 data collection period, the interquartile range (IQR) of performance scores ranged from 42% to 78%. During the October 2013 through September 2014 data collection period, the IQR ranged from 54% to 83%. While median performance is improving, there is an ongoing opportunity for improvement in facility performance.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
Data have been included in Section 1b.2; these data represent national performance over time, from October 2012 through September 2014 data collection periods.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Using 2014 data submitted to the Clinical Data Warehouse (CDW), we evaluated the effect of patient and facility characteristics on the likelihood of each patient having a head CT or MRI scan interpreted within 45 minutes of ED arrival for patients with a principal diagnosis associated with acute ischemic or hemorrhagic stroke, who arrived at the ED within two hours of the time last known well, and who had an order for a head CT or MRI scan. Using a logistic regression model, we assessed the impact of patient and facility characteristics for the 28,236 patients who met these criteria.

Primary results from the regression were related to patient demographics. African-American patients were less likely than White patients to have a head CT or MRI scan interpreted within 45 minutes of ED arrival (OR= 0.865, p=0.003). In comparison to non-Hispanic patients, Hispanic patients were less likely to have a head CT or MRI scan interpreted within 45 minutes of ED arrival (OR= 0.80, p=0.010). Finally, female patients were less likely than male patients to have a head CT or MRI scan interpreted within 45 minutes of ED arrival (OR= 0.86, p<0.001).

Facility characteristics also played a role in determining whether a patient had a head CT or MRI scan interpreted within 45 minutes of ED arrival for patients with a principal diagnosis associated with acute ischemic or hemorrhagic stroke, who arrived at the ED within two hours of the time last known well, and who had an order for a head CT or MRI scan. When compared to patients treated in facilities with fewer than 50 beds (a proxy for facility size), patients treated in facilities with 51-100 beds (OR= 1.45, p<0.001), 101-250 beds (OR= 2.12, p<0.001), 251-500 beds (OR=1.81, p<0.001), and 500 or more beds (OR 1.143, p=0.007) were more likely to have a head CT or MRI scan interpreted within 45 minutes of ED arrival. Patients treated in a major teaching facility were less likely than those treated in a non-teaching facility to have a head CT or MRI scan interpreted within 45 minutes of ED arrival (OR= 0.62, p<0.001).

While the regression model identified subpopulations of patients and facilities for which there are statistically significant differences in the likelihood of a patient having a head CT or MRI scan interpreted within 45 minutes of ED arrival for patients with a principal diagnosis associated with acute ischemic or hemorrhagic stroke, who arrived at the ED within two hours of the time last known well, and who had an order for a head CT or MRI scan, these disparities do not indicate a need for adjustment of the measure specifications. Adjusting for these differences would mask underlying differences in quality of care. As this is a process measure, there should be no difference in the standard of care for these patients; we believe these statistically significant differences are driven by variation in provider practice. Consequently, we do not believe risk adjustment or stratification is necessary or appropriate for this measure.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
Not applicable as there is available data for disparities in care.

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare
Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality
1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.
According to the Centers for Disease Control and Prevention (CDC), stroke is the fifth leading cause of death (Kochanek et al. 2014) and is a leading cause of serious long-term disability. Nearly 800,000 people experience a new or recurrent stroke in the United States every year (Mozaffarian et al. 2015). Ischemic strokes account for approximately 87% of strokes, and tPA is the only FDA-approved treatment for ischemic stroke (Spiotta et al. 2015). The FDA recommends that tPA be administered as soon as possible but within three hours after the stroke symptoms begin. In a study of 71,169 patients at 1,030 hospitals participating in the Get With the Guidelines-Stroke (GWTG-Stroke) program found that improved timeliness of tPA administration was associated with lower in-hospital mortality, decreased intracranial hemorrhage risk, and an increase in the percentage of patients discharged home (Fonarow et al. 2014).

Despite clear evidence that timely administration of tPA is associated with improved health outcomes, multiple studies indicate that fewer than one-third of stroke patients arrive at the ED within the recommended time frame (within two hours of symptom onset) to be eligible for tPA (Mozaffarian et al. 2015; Pittenger et al. 2014). While symptom-to-arrival time is beyond the control of facilities, it is imperative for facilities to perform head imaging to promptly evaluate potentially eligible patients to initiate treatment as soon as possible.

To complement a review of the literature, we extracted national statistics on ED use from the Healthcare Cost and Utilization Project (HCUP) National Emergency Department Sample (NEDS) to estimate the total patient population of patients with a primary diagnosis associated with acute ischemic or hemorrhagic stroke. In 2012, there were an estimated 539,189 ED visits with a primary diagnosis of stroke. Most patients were ages 65-84 (45.2%), followed by patients ages 45-64 (30.4%), 85+ (18.7%), 18-44 (5.4%), and finally patients under 18 (0.3%). There were slightly more females patients (51.6%) than male patients (48.4%).

Additionally, the Department of Health and Human Services (DHHS) includes stroke care in multiple national campaigns such as the Million Hearts Campaign and Healthy People 2020. Both of these initiatives galvanize existing efforts and new programs to improve cardiovascular health and quality of life through prevention, detection, and treatment of stroke and heart attacks (Frieden and Berwick 2011; DHHS 2014). Reducing door-to-interpretation time through the continued reporting of NQF #0661 can help DHHS achieve its Healthy People 2020 objective of increasing the proportion of eligible patients with stroke who receive acute reperfusion therapy within three hours from symptom onset (DHHS 2014).

The literature and utilization data demonstrate that acute stroke treatment remains a high priority aspect of healthcare because stroke is a leading cause of morbidity/mortality that affects a large number of individuals with severe consequences for poor quality of care.

1c.4. Citations for data demonstrating high priority provided in 1a.3
1.) Cheng NT, Kim AS. Intravenous Thrombolysis for Acute Ischemic Stroke Within 3 Hours Versus Between 3 and 4.5 Hours of Symptom Onset. Demaerschalk BM, ed. The Neurohospitalist. 2015;5(3):101-109. doi:10.1177/1941874415583116.
2.) DHHS. 2014. HDS-19.3 (Developmental) Increase the proportion of eligible patients with strokes who receive acute reperfusion therapy within 3 hours from symptom onset. http://www.healthypeople.gov/node/4581/data_details.
3.) Fonarow GC, Zhao X, Smith EE, Saver JL, et al. Door-to-needle times for tissue plasminogen activator administration and clinical outcomes in acute ischemic stroke before and after a quality improvement initiative. JAMA. 2014;311(16):1632-1640. doi:10.1001/jama.2014.3203.
4.) Frieden TR and Berwick D. The "Million Hearts" initiative – preventing heart attacks and strokes. N Engl J Med. 2011;365:e27.
5.) Jauch EC, Saver JL, Adams HP Jr, Bruno A, Connors JJ, Demaerschalk BM, Khatri P, McMullan PW Jr, Qureshi AI, Rosenfield K, Scott PA, Summers DR, Wang DZ, Wintermark M, Yonas H; on behalf of the American Heart Association Stroke Council, Council on Cardiovascular Nursing, Council on Peripheral Vascular Disease, and Council on Clinical Cardiology. Guidelines for the early

management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44. Guideline available at: http://stroke.ahajournals.org/content/early/2013/01/31/STR.0b013e318284056a.full.pdf+html
6.) Kochanek KD, Murphy SL, Xu JQ, Arias E. Mortality in the United States, 2013. NCHS data brief, no 178. Hyattsville, MD: National Center for Health Statistics. 2014.
7.) Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després J-P, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Liu S, Mackey RH, Matchar DB, McGuire DK, Mohler ER 3rd, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Nichol G, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner MB; on behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circulation. 2015;131:eXX–eXXX.
8.) Pittenger J, Fuhrman S, Hansen A, et al. Abstract 268: Evaluating pre-hospital response when stroke symptoms are present through hospital arrival mode and last known well to arrival times. Circulation: Cardiovascular Quality and Outcomes. 2014;7:A268.
9.) Saver JL, Fonarow GC, Smith EE, et al. Time to treatment with intravenous tissue plasminogen activator and outcome from acute ischemic stroke. JAMA. 2013;309(23):2480-8. doi: 10.1001/jama.2013.6959.
10.) Spiotta AM, Chaudry MI, Hui FK, et al. Evolution of thrombectomy approaches and devices for acute stroke: a technical review. J NeuroIntervent Surg. 2015;7:2-7. doi:10.1136/neurintsurg-2013-011022.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
This measure is not a PRO-PM measure.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology, Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier2&cid=1196289981244

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** NQF_0661_Measure_Code_Set.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
NQF #0661 was granted time-limited endorsement in January 2011. The Consensus Standards Approval Committee (CSAC) removed the time-limited endorsement after reviewing testing results for NQF #0661 in April 2014, moving the measure to a fully endorsed status. Since 2011, the measure specifications have been updated to reflect clinical changes in appropriate stroke diagnosis or ED evaluation and management (E/M) codes; to address stakeholder feedback; or to harmonize with the Hospital Inpatient Quality Reporting (HIQR) program. In 2012, the measure algorithm was updated to reject cases with less than zero measurement values and to exclude outliers from being included in the data set. The Discharge Status data element was changed to Discharge Code. To facilitate abstraction, suggested data sources and a hierarchy of provider types were added for the Time Last Known Well, Date Last Known Well, and Last Known Well data elements.

In 2015, as part of the annual measure maintenance and review process, all ICD-9-CM diagnosis codes were updated to corresponding ICD-10-CM diagnosis codes. The data accuracy section of the Measure Information Form was updated with a disclaimer that there may be variation in the assignment of ICD-10-CM codes by provider, facility, and documentation protocol for the chart-abstracted data elements The proposed updates were supported by independent reviews by the experts supporting the HIQR program, which has a related stroke measure (STK-4: Thrombolytic Therapy, NQF #0437).

The Notes for Abstraction for key data elements were updated to add examples and clarifying language to address stakeholder feedback and to better align with NQF #0437; affected data elements include: Arrival Time and Head CT or MRI Scan Interpretation Date.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The number of acute ischemic stroke or hemorrhagic stroke patients who arrive at the ED within two hours of the onset of symptoms and have a head CT or MRI scan interpreted within 45 minutes of ED arrival.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator: Head CT or MRI scan interpretations within a three month period, aggregated on a rolling basis for acute ischemic or hemorrhagic stroke patients who arrive to the ED within 45 minutes of symptom onset
Denominator: Head CT or MRI scan orders within a three month period, aggregated on a rolling basis for acute ischemic or hemorrhagic stroke patients who arrive to the ED within 45 minutes of symptom onset

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The numerator is defined by six evaluation and management (E/M) codes and 102 ICD-10-CM diagnosis codes included in the code set for this measure; these detailed lists can be found in the Excel workbook provided for Section S2b.

The numerator includes patients age 18 or older who were last known well within two hours of ED arrival and had a head CT or MRI ordered and interpreted within 45 minutes of ED arrival. Numerator exceptions include:

- Date Last Known Well is equal to UTD
- Time Last Known Well is equal to UTD
- Arrival Time is equal to UTD
- Head CT Scan or MRI Interpretation Date is equal to UTD
- Head CT Scan or MRI Interpretation Date is equal to UTD

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The number of acute ischemic stroke or hemorrhagic stroke patients who arrive at the ED within two hours of the onset of symptoms and have a head CT or MRI scan ordered.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The denominator is defined by six evaluation and management (E/M) codes and 104 ICD-10-CM diagnosis codes included in the code set for this measure; these detailed lists can be found in the Excel workbook provided for Section S2b.

The denominator includes patients age 18 or older who were last known well within two hours of ED arrival and had a head CT or MRI ordered.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*

Studies are excluded for any patients under 18 years of age, patients who expired in the ED, or patients who left the ED against medical advice or discontinued care. Additionally, patients who do not arrive to the ED within two hours of symptom onset or who do not have a head CT or MRI scan ordered are excluded from the target population.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Studies are excluded for any patients that meet any of the following criteria:

- Patients less than 18 years of age
- Patients who expired (discharge code = 6)
- Patients who left the emergency department against medical advice or discontinued care (discharge code = 7 or 8)
- Patients who have a head CT or MRI scan order equal to "No"
- Patients who have a Last Known Well field equal to "No"

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not applicable; this measure does not stratify its results.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not applicable; this measure does not risk adjust.

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Provided in response box S.15a

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
No risk model specifications are provided, as risk adjustment or stratification is not necessary for this measure.

S.16. **Type of score:**
Other (specify):
If other: Percentage

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*

This measure calculates the percentage of acute ischemic stroke or hemorrhagic stroke patients who arrive at the ED within two hours of the onset of symptoms and have a head CT or MRI interpreted within 45 minutes of ED arrival. The measure is calculated based on four consecutive quarters of hospital outpatient claims data, as follows:

1. Check E/M Code; if on Table 1.0 (in the Excel workbook provided for Section S2b), proceed
2. Calculate Patient Age (Outpatient Encounter Date - Birthdate)
3. Check Patient Age; if >= 18, proceed
4. Check ICD-10-CM Principal Diagnosis Code; if on Table 8.0 (in the Excel workbook provided for Section S2b), proceed
5. Check Discharge Code; exclude any patients with code 6, 7, or 8
6. Check Head CT or MRI Scan Order; if "Yes," proceed
7. Check Last Known Well; if "Yes," proceed
8. Check Date Last Known Well; if a Non-Unable to Determine (UTD) value, proceed
9. Check Time Last Known Well; if a Non-UTD value, proceed
10. Check Arrival Time; if a Non-UTD value, proceed
11. Calculate measurement value (Arrival Time minus Time Last Known Well)
12. Check measurement value; if >= 0 min and <= 120 min, record as the denominator and proceed
13. Check Head CT or MRI Scan Interpretation Date; if a Non-Unable to Determine (UTD) value, proceed
14. Check Head CT or MRI Scan Interpretation Time; if a Non-Unable to Determine (UTD) value, proceed
15. Calculate measurement value (Arrival Time minus Head CT or MRI Scan Interpretation Time)
16. Check measurement value; if >= 0 min and <= 45 min, record as the numerator
17. Aggregate denominator and numerator counts by Medicare provider number
18. Measure = numerator counts / denominator counts [The value should be recorded as a percentage]

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Sampling is a process of selecting a representative part of a population in order to estimate the hospital's performance without collecting data for its entire population. Using a statistically valid sample, a hospital can measure its performance in an effective and efficient manner. Sampling is a particularly useful technique for performance measures that require primary data collection from a source such as the medical record. Sampling should not be used unless the hospital has a large number of cases in the outpatient population because a fairly large number of sample cases are needed to achieve a representative sample of the population. For the purpose of sampling outpatient department quality measures, the terms "sample," "effective sample," and "case" are defined below:
• The "sample" is the fraction of the population that is selected for further study.
• "Effective sample" refers to the part of the sample that makes it into the denominator of an outpatient measure set. This is defined as the sample for an outpatient measure set minus all the exclusions and contraindications for the outpatient measure set in the sample.
• A "case" refers to a single record (or an encounter) within the population. For example, during the first quarter a hospital may have 100 patients who had a principal diagnosis associated with the OP-1, 2, 3, 4, and 5 measures. The hospital's outpatient population would include 100 cases or 100 outpatient records for these measures during the first quarter.

To obtain statistically valid sample data, the sample size should be carefully determined, and the sample cases should be randomly selected in such a way that the individual cases in the population have an equal chance of

being selected. Only when the sample data truly represent the whole population can the sample-based performance outpatient measure set data be meaningful and useful. Each hospital is ultimately responsible for adhering to the sampling requirements outlined in this manual.

As a general rule/policy of CMS, providers are encouraged to submit as many cases as possible up to the entire population of cases if reasonably feasible. For example, if the raw data can be easily extracted from an existing electronic database or the abstraction burden is manageable, providers should consider submitting the entire population of cases that meet the initial selection criteria. Otherwise, a statistically valid sample can be selected.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
This measure does not use survey data.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
The measure does not make any adjustments for missing data. If data are missing, the case will proceed to Measure Category Assignment of X and will be rejected. While abstractors cannot submit missing data, they submit a value of "UTD" for select data elements. Depending on the data element the case is then either excluded from the denominator or excepted from the numerator. Frequency and distribution of data with a value of "UTD" are reported in the attached Measure Testing form.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
An electronic data collection tool is made available from vendors or facilities can download the free CMS Abstraction & Reporting Tool (CART). Paper tools for manual abstraction, which are posted on www.QualityNet.org, are also available for the CART tool. These tools are posted on www.QualityNet.org.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Emergency Medical Services/Ambulance, Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable; this is not a composite measure.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF_0661_Measure_Testing_Form.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 0661
Measure Title: Head CT or MRI Scan Results for Acute Ischemic Stroke or Hemorrhagic Stroke Patients who Received Head CT or MRI Scan within 45 Minutes of ED Arrival
Date of Submission: 1/15/20161/15/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance

score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of

missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: **(*must be consistent with data sources entered in S.23*)**	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☒ administrative claims	☒ administrative claims

☐ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

a) Datasets used to define the initial patient population:

- The initial patient population is identified using chart-abstracted data for a sample of ED encounters with at least one of the following Current Procedural Terminology (CPT) codes for evaluation and management (E/M): 99281, 99282, 99287, 99284, 99285, or 99291. The initial patient population includes cases for patients 18 years and older, as of the date of the encounter, with a principle diagnosis associated with an acute ischemic or hemorrhagic stroke, identified by using any of the following International Classification of Diseases version 9 (ICD-9) codes: 430, 431, 433.01, 433.11, 433.21, 433.31, 433.81, 433.91, 434.01, 434.11, 434.91, or 436.

b) Datasets used to define the denominator:

- The denominator is identified using chart-abstracted data for a sample of cases for patients included in the initial patient population.

c) Datasets used to identify denominator exclusions:

- Denominator exclusions are identified using chart-abstracted data of cases for patients included in the denominator. Denominator exclusions capture cases for patients where any of the following conditions are met:
    - *Discharge Code* is equal or equivalent to "Expired," "Left Against Medical Advice/AMA" or "not documented or unable to determine (UTD)"
    - *Head CT or MRI Scan Order* is equal to missing or "No"
    - *Last Known Well* is equal to "No"
    - *Time Last Known Well* is greater than 120 minutes

d) Datasets used to capture the numerator:

- The numerator is identified using chart-abstracted data of cases for patients included in the denominator. The numerator includes cases for patients where either of the following conditions are met:
    - *ED Arrival Time* to *Head CT Scan Interpretation Time* is within 45 minutes
    - *ED Arrival Time* to *MRI Scan Interpretation Time* is within 45 minutes

e) Datasets used to identify numerator exceptions:

- Numerator exceptions are identified by using chart-abstracted data of cases for patients included in the denominator. Numerator exceptions include cases of patients for whom any of the following conditions are met:
    - *Date Last Known Well* is equal to "UTD" –*Time Last Known Well* is equal to "UTD,"
    - *Arrival Time* is equal to "UTD"
    - *Head CT or MRI Scan Interpretation Date* is equal to "UTD"
    - *Head CT or MRI Scan Interpretation Time* is equal to "UTD"

1.3. What are the dates of the data used in testing? January 1, 2014–December 31, 2014January 1, 2014–December 31, 2014

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☒ other: nationalnational	☒ other: nationalnational

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The number of measured entities (hospital emergency departments) varies by testing type; see Section **1.7** for details.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

The number of patients varies by testing type; see Section **1.7** for details.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Reliability Testing:
Data Source: Hospital Compare downloadable file [maintained by the Centers for Medicare & Medicaid Services (CMS)]
Dates: Denominator: January 1, 2014-December 31, 2014; Numerator: January 1, 2014-December 31, 2014; Exclusions: January 1, 2014-December 31, 2014; Exceptions: January 1, 2014-December 31, 2014
Number of Facilities: 958
Denominator Cases: 17,162
Numerator Cases: 11,741
Level of Analysis: Facility
Patient Characteristics: Not applicable

Validity Testing—*Empirical Validity*
Data Source: Validation mismatches: Hospital Outpatient Quality Reporting (Hospital OQR) Clinical Data Warehouse (CDW)
Dates: April 1, 2014-March 31, 2015
Sampled Population: 774
Level of Analysis: Data element
Patient Characteristics: Not applicable

Validity Testing – *Face Validity*
Data Source: Structured qualitative survey completed by the stroke and acute myocardial infarction expert work group (EWG) members
Date Collected: October - November 2015
Number of Responses: 5
Respondent Characteristics: Respondents were asked to self-identify as one or more of the following categories: clinician (4); healthcare administration (1); management (1), other - payer consultant (1), other - Institute Director, Association (1).

Exclusions Analysis
Data Source: Denominator: CDW; Numerator: CDW; Exclusions: CDW
Dates: Denominator: January 1, 2014-December 31, 2014; Numerator: January 1, 2014-December 31, 2014; Exclusions: January 1, 2014-December 31, 2014; Exceptions: January 1, 2014-December 31, 2014
Number of Facilities: 3,614
Sampled Population: 105,898
Denominator Cases: 28,236
Numerator Cases: 18,480
Level of Analysis: Case
Denominator Patient Characteristics: Gender (% Male): 49.0; Mean Age (Years): 66.8 (St. Dev.: 15.3); Race (% Minority): 19.3

Risk Adjustment/Stratification
N/A- No risk adjustment or stratification was performed.

Identification of Statistically Significant & Meaningful Differences in Performance
Data Source: Hospital Compare downloadable file
Dates: Denominator: January 1, 2014-December 31, 2014; Numerator: January 1, 2014-December 31, 2014; Exclusions: January 1, 2014-December 31, 2014; Exceptions: January 1, 2014-December 31, 2014
Number of Facilities: 958
Denominator Cases: 17,162
Numerator Cases: 11,741
Level of Analysis: Facility
Patient Characteristics: Not applicable

Comparability of Performance Scores when more than one Set of Specifications
N/A- This measure only uses one set of specifications.

Missing Data Analysis and Minimizing Bias
Data Source: Denominator: CDW; Numerator: CDW; Exclusions: CDW
Dates: Denominator: January 1, 2014-December 31, 2014; Numerator: January 1, 2014-December 31, 2014; Exclusions: January 1, 2014-December 31, 2014; Exceptions: January 1, 2014-December 31, 2014
Number of Facilities: 3,614
Sampled Population: 105,898
Denominator Cases: 28,236
Numerator Cases: 18,480
Level of Analysis: Case
Denominator Patient Characteristics: Gender (% Male): 49.0; Mean Age (Years): 66.8 (St. Dev.: 15.3); Race (% Minority): 19.3

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

We assessed patient-level SDS factors as part of the regression model reported in Section **1b.4**, which provides an overview of disparities in care for patient sub-populations. We based this analysis on SDS variables included in the CDW data:

- Age
- Sex
- Race
- Ethnicity

While an analysis of SDS factors is important in understanding differences in care for patient sub-populations, this measure is a process measure that is neither risk-adjusted nor risk-stratified. We determined that risk adjustment and risk stratification were not appropriate based on the

current evidence base and the measure construct. Additional information on this determination is provided in Section **2b4.2**.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

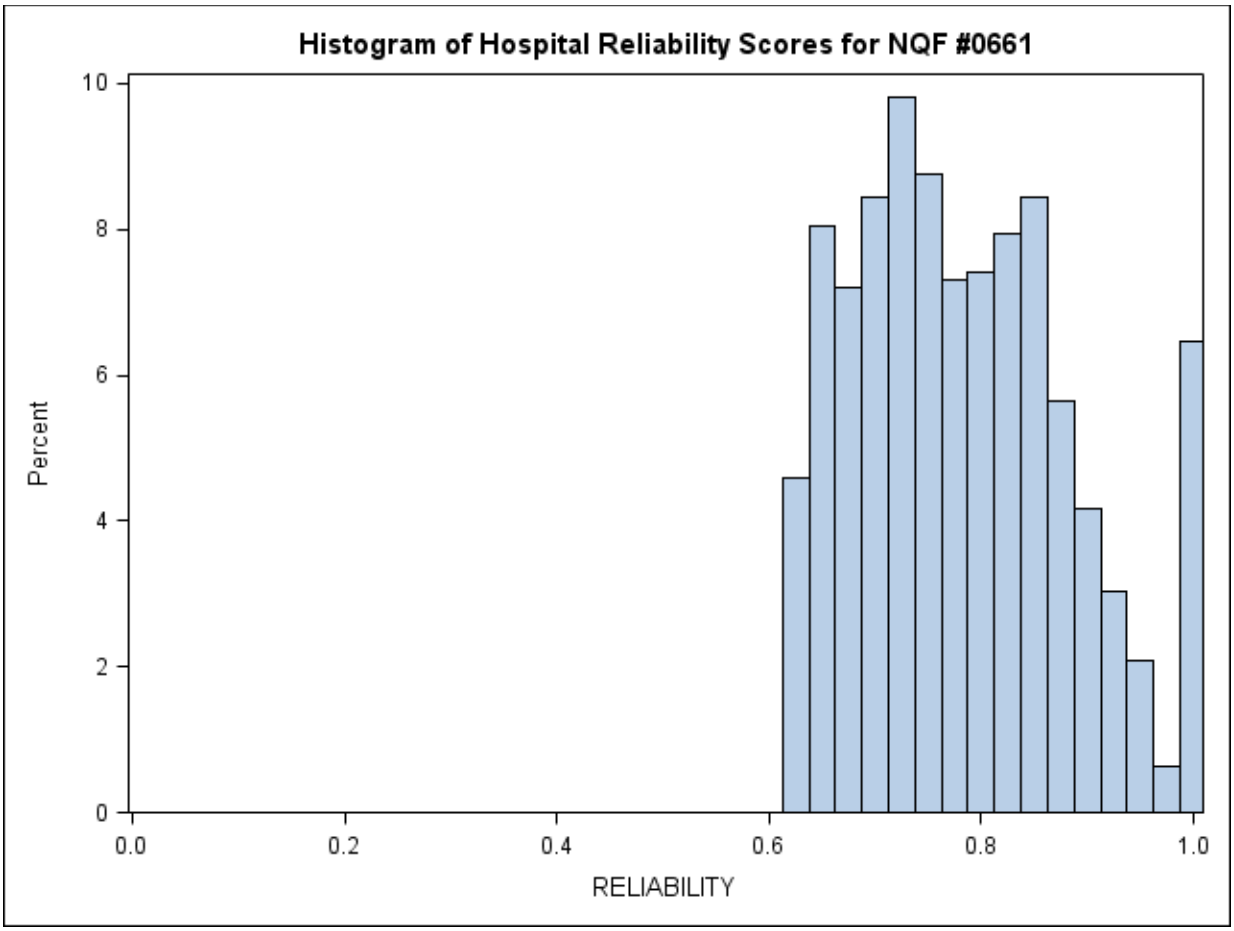
Reliability was calculated in accordance with the methods discussed in *The Reliability of Provider Profiling: A Tutorial* (2009). This approach calculates the ability of the measure to distinguish between the performances of different facilities. Specifically, the testing calculated the signal-to-noise ratio for each facility meeting the minimum case count, established by the measure calculation contractor, during the 2014 data collection period, with higher scores indicating greater reliability. The reliability score is estimated using a beta-binomial model, which is appropriate for the reliability testing of pass/fail measures. The reliability score for each facility is a function of the facility's sample size and score on the measure, and the variance across facilities.

REFERENCES:

Adams JL. The reliability of provider profiling: a tutorial. Santa Monica, CA: RAND Corporation. 2009. Retrieved from http://www.rand.org/pubs/technical_reports/TR653.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Figure 1 (below) is a histogram of the distribution of the reliability scores for the facilities meeting the minimum case count requirements during the 2014 data collection period. Reliability scores ranged from 0.62 to 1.00, with a median reliability score of 0.77. Higher scores denote greater reliability.



2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e.*, *what do the results mean and what are the norms for the test conducted?*)

Calculated using a beta-binomial model, a median reliability score of 0.77 is indicative of strong measure reliability. The results of this test indicate that the measure is able to identify true differences in performance between individual facilities.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The validity of the measure using both qualitative and quantitative analyses was assessed. Empirical validity of the measure performance score was assessed by calculating a kappa statistic assessing the agreement between facility abstraction and auditor (CDAC) abstraction. Face validity of the measure score was systematically assessed through survey of the EWG.

The kappa statistic is the test statistic used to measure interrater reliability and demonstrates the percent agreement between two sources for the same observation. Kappa values range from 0.00 to 1.00, where a value of 0.00 indicates zero agreement between two sources and a value of 1.00 indicates complete agreement between two sources. To estimate the statistical significance associated with the kappa statistic, p-values can be calculated. P-values of less than 0.001 suggest very high levels of statistical significance, and suggest the results are not due to chance. For NQF measure #0661, the kappa statistic was used to estimate the level of agreement between the facility's abstraction of numerator and denominator cases versus CDAC's abstraction of the same sample of cases. The p-values associated with these estimates are also reported.

Landis & Koch, 1977 offer the following classification of kappa interpretation:

<0	Poor agreement
0.00–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–1.00	Almost perfect agreement

REFERENCES:

Landis, J. & Koch, G. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174. 1977.

Face validity of the measure score was systematically assessed through survey of the EWG. Five EWG members participated in the data collection. Respondent perspectives include clinicians, management, and healthcare administration. Prior to responding to questions related to measure-score face validity, EWG members were provided detailed measure specifications.

The following questions and statements related to measure-score face validity were posed to the EWG:

1. Patients who have a head CT scan or MRI ordered and interpreted within 45 minutes of ED arrival can be accurately captured using chart-abstracted data.
2. The measure successfully assesses the timely interpretation of head imaging for acute ischemic and hemorrhagic stroke patients.

Responses to questions *1* and *2* in the measure-score face-validity section were collected using a five-point Likert scale: strongly agree, agree, undecided, disagree, strongly disagree, and do not know.

The face validity assessment was complemented by an empirical evaluation of the eight critical data elements, calculating a rate of agreement between facility abstraction and auditor (CDAC) abstraction. The empirical analysis used data element values for 774 cases abstracted by CDAC, which were previously abstracted by facilities. The data was collected from April 1, 2014-March 31, 2015.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Results of empirical validity testing indicate moderate to complete levels of agreement between the facility's abstraction of numerator and denominator cases versus CDAC's abstraction of the same sample of cases. The kappa statistic for numerator cases equals 0.52, with a p-value of less than 0.001. The kappa statistic for denominator cases equals 1.00, with a p-value of less than 0.001.

Results of the face-validity assessment indicate that a diverse group of stakeholders support the validity of the measure. Results for each of the questions provided above follow.

1. Patients who have a head CT scan or MRI ordered and interpreted within 45 minutes of ED arrival can be accurately captured using chart-abstracted data.

Response Option	Response Percentage	Response Count
Strongly Agree	60.0%	3
Agree	40.0%	2
Undecided	0.0%	0
Disagree	0.0%	0
Strongly Disagree	0.0%	0
Do Not Know or Not Applicable	0.0%	0

2. *The measure successfully assesses the timely interpretation of head imaging for acute ischemic and hemorrhagic stroke patients.*

Response Option	Response Percentage	Response Count
Strongly Agree	40.0%	2
Agree	60.0%	3
Undecided	0.0%	0
Disagree	0.0%	0
Strongly Disagree	0.0%	0
Do Not Know or Not Applicable	0.0%	0

These findings are further supported by the results of empirical validity testing, which indicate moderate to strong levels of agreement between the facility's abstraction of data elements versus CDAC's abstraction of data elements for the same sample of cases. The rate of agreement, by data element, ranged from 52.7% to 98.4%. The rate of agreement was strong for dichotomous variables, as well as those based on administrative data. This included variables, such as *Scan Ordered*, *Discharge Code*, and *Scan Order Date*. Agreement was moderate for clinical variables related to time, such as *Time Last Known Well* and *Scan Order Time*.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e.*, *what do the results mean and what are the norms for the test conducted?*)

Results of the quantitative and qualitative analysis are positive and support the validity of the measure and its calculation. Based on the Landis and Koch classification scale, described in section **2b2.2**, there was moderate agreement between facility and auditor abstraction of numerator cases (kappa=0.52; p-value <0.001) and complete agreement between facility and auditor abstraction of denominator cases (kappa= 1.00; p-value <0.001). This suggests strong validity for the measure, as currently specified.

The EWG, composed of five stakeholders representing healthcare administration, management, payer consultants, associations, and clinicians with expertise in cardiology, neuro-radiology, emergency medicine, and emergency nursing, provided feedback on the face validity of NQF #0661 through an online survey. All members agreed or strongly agreed that patients who have a head CT scan or MRI ordered and interpreted within 45 minutes of ED arrival can be accurately captured using chart-abstracted data. Similarly, they agreed or strongly agreed that NQF #0661 successfully assesses the timely interpretation of head imaging for acute ischemic and hemorrhagic stroke patients. The respondents generally support the face validity of NQF #0661.

The rate of agreement between facility and CDAC abstraction ranged from moderate to strong across the data elements used to calculate OP-23. There is not a benchmark for interpretation of rates of agreement for chart-abstracted measures; however, reported results are in line with testing results from measure development (previously reviewed by NQF in 2012), as well as feedback from a panel of experts.

2b3. EXCLUSIONS ANALYSIS

NA □ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

We tested measure exclusions and numerator exceptions to determine the prevalence of each exclusion and exception, by facility, and at an aggregate level. The analysis tested measure exclusions and numerator exceptions during the 2014 data collection period. Measure exclusions include all cases meeting one or more criteria listed in Section **1.2c**, above. Numerator exceptions include cases meeting one or more criteria listed in Section **1.2d**, above. To

supplement the empirical results, we systematically assessed the face validity of current exclusions through survey of the EWG based on responses from five EWG members.

The face validity of exclusions was assessed, using the following questions and statements:

1. To be included in the measure population, each patient must receive care in the ED for a stroke. These patients are identified based on ICD-9 principal diagnosis codes and E&M codes. From this initial patient population, certain patients are excluded from NQF #0661 based on the situations listed in the table below.[1] Please evaluate the appropriateness of each of the CURRENT exclusion criteria.
2. For NQF #0661, do you foresee any challenges in capturing any of the exclusions using chart-abstracted data?

Responses to question *1* were collected using keep/remove response options; responses to question *2* were collected using yes/no response options.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

We examined overall frequencies and proportions of cases excluded for each exclusion/exception criterion, among all sampled cases, for 3,614 facilities submitting eligible cases in 2014. The sampled population included 105,898 cases where a patient (age 18 years or older) presented with an ischemic or hemorrhagic stroke to an ED.

Overall Occurrence and Distribution across Facilities for Measure Exclusions and Exceptions

Data Element	Denominator Exclusion or Numerator Exception?		Overall Occurrence		Distribution Across Facilities (%)		
	Denominator Exclusion	Numerator Exception	N	%	25th	50th	75th
Discharge Code	X		4,627	4.4	0.0	2.5	6.7
Head CT Scan or MRI Order	X		3,537	3.3	0.0	1.6	5.3
Last Known Well	X		44,328	41.9	28.6	41.4	52.6
Date Last Known Well		X	100	0.1	0.0	0.0	0.0
Time Last Known Well		X	1,506	1.4	0.0	0.0	0.0

[1] Respondents were provided a table detailing the key measure exclusions.

Data Element	Denominator Exclusion or Numerator Exception?		Overall Occurrence		Distribution Across Facilities (%)		
	Denominator Exclusion	Numerator Exception	N	%	25th	50th	75th
Arrival Time		X	5	0.0	0.0	0.0	0.0
Last Known Well Minutes	X		25,170	23.8	14.3	22.2	30.4
Head CT Scan or MRI Interpretation Date		X	87	0.1	0.0	0.0	0.0
Head CT Scan or MRI Interpretation Time		X	373	0.4	0.0	0.0	0.0
Total Denominator Exclusions	4 exclusions	-	77,662	73.40	-	-	-
Total Numerator Exceptions	-	5 exceptions	2,071	2.00	-	-	-
Total Removed from the Denominator or Numerator	9 exceptions and exclusions		79,733	75.40	-	-	-

As indicated in the table above, 73.4% of the initial patient population is excluded from the denominator. This may be partially explained by clinical factors, such as stroke patients presenting to the ED after two hours of symptom onset, or by limitations of chart abstracted documentation, which might have a large volume of missing and/or abnormal cases. However, the use of minimum case counts ensures that we report performance scores for facilities that have an adequate number of cases after the application of these exclusions.

The EWG provided feedback through an online survey on the appropriateness of the three categories of excluded populations: patients less than 18 years of age, patients who expired before discharge, and patients who left against medical advice (AMA) or discontinued care. The survey results indicate that a diverse group of stakeholders generally support the exclusions for NQF #0661, and do not foresee significant challenges in capturing the exclusions described above using chart-abstracted data.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**,*

the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion)

As seen in the table reported in Section **2b3.2** above, the frequency of exclusions/exceptions varied substantially across facilities. Measure exclusion and numerator exception criteria are in alignment with clinical guidelines and also ensure that all cases included in the measure have sufficient denominator and numerator information to calculate the performance score. After identification of cases for patients 18 years and older with a principal diagnosis associated with acute ischemic or hemorrhagic stroke, measure exclusion and numerator exception criteria are applied:

- *Discharge Code* is a measure exclusion criterion. Cases for patients where *Discharge Code* equals "Expired", "Left Against Medical Advice/AMA", or "UTD" are excluded from the measure denominator. Overall, 4.4% of cases for patients included in the initial patient population are excluded from the denominator based on *Discharge Code*. There is notable variability in the proportion of cases excluded based on *Discharge Code* values across facilities, with an inter-quartile range of 0.0% to 6.7%.

- *Head CT Scan or MRI Order* is a measure exclusion criterion. Cases for patients where *Head CT Scan or MRI Order* equals "No" are excluded from the denominator. This criterion is based off the fact that the numerator is dependent upon a head CT or MRI scan being performed. Overall, 3.3% of cases included in the initial patient population, were excluded from the measure denominator based on *Head CT Scan or MRI Order*. There is notable variability in the proportion of excluded *Head CT Scan or MRI Order* values across facilities, with an interquartile range from 0.0% to 5.3%.

- *Last Known Well* is a measure exclusion criterion. It is a binary variable that indicates if there are values for both *Date Last Known Well* and *Time Last Known Well.* Cases for patients where *Last Known Well* is equal to "No" are excluded from the measure denominator. Overall, 41.9% of cases for patients included in the initial patient population had a *Last Known Well* value equal to "No." There is large variability in the proportion of excluded *Last Known Well* values across facilities, with an interquartile range from 28.6% to 52.6%.

- *Date Last Known Well* is a numerator exception criterion. If *Date Last Known Well* is equal to "UTD," the case is not included in the measure numerator but remains in the measure denominator. Overall, 0.1% of patients included in the denominator have a "UTD" value for *Date Last Known Well*. While there is limited variability in the proportion of excepted cases across facilities, the exception remains important as a "UTD" value for this data element makes it impossible to determine whether the patient received timely interpretation of their head CT or MRI.

- *Time Last Known Well* is a numerator exception criterion. If *Time Last Known Well* is equal to "UTD," the case is not included in the measure numerator but remains in the measure denominator. Overall, 1.4% of patients included in the denominator have a "UTD" value for *Time Last Known Well*. While there is limited variability in the

proportion of excepted cases across facilities, the exception remains important as a "UTD" value for this data element makes it impossible to determine whether the patient received timely interpretation of their head CT or MRI.

- *Arrival Time* is a numerator exception criterion. If *Arrival Time* is equal to "UTD," the case is not included in the measure numerator but remains in the measure denominator. Overall, less than 0.1% of patients included in the denominator have a "UTD" value for *Arrival Time*. While there is limited variability in the proportion of excepted cases across facilities, the exception remains important as a "UTD" value for this data element makes it impossible to determine whether the patient received timely interpretation of their head CT or MRI.

- *Last Known Well Minutes* is a measure exclusion criterion. It is a measurement value calculated from *Arrival Time* and *Time Last Known Well.* If *Last Known Well Minutes* is greater than 120 minutes, the case is excluded from the measure. This criterion is based off of clinical guidelines for the most appropriate time window to identify and treat acute stroke. Overall, 23.8% of patients eligible for the measure have a value greater than 120 minutes for *Time Last Known Well Minutes; although this value may appear high, multiple studies have found that fewer than one-third of stroke patients arrive at the ED within two hours of symptom onset (Mozaffarian et al. 2015; Pittenger et al. 2014)*. There is variability in the proportion of excluded values for *Last Known Well Minutes* across facilities, with an interquartile range of 14.3% to 30.4%.

- *Head CT Scan or MRI Interpretation Date* is a numerator exception criterion. If *Head CT Scan or MRI Interpretation Date* is equal to "UTD," the case is not included in the measure numerator but remains in the measure denominator. Overall, 0.1% of patients included in the denominator have a "UTD" value for *Head CT Scan or MRI Interpretation Date*. While there is limited variability in the proportion of excepted cases across facilities, the exception remains important as a "UTD" value for this data element makes it impossible to determine whether the patient received timely interpretation of their head CT or MRI.

- *Head CT Scan or MRI Interpretation Time* is a numerator exception criterion. If *Head CT Scan or MRI Interpretation Time* is equal to "UTD," the case is not included in the measure numerator but remains in the measure denominator. Overall, 0.4% of patients included in the denominator have a "UTD" value for *Head CT Scan or MRI Interpretation Date*. While there is limited variability in the proportion of excepted cases across facilities, the exception remains important as a "UTD" value for this data element makes it impossible to determine whether the patient received timely interpretation of their head CT or MRI.

Results of the survey of the EWG also support the face validity of the exclusions and exceptions for NQF #0661, and indicate that these exclusions are consistent with prevailing gold standards of care or are necessary to support measure calculation.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

This measure is a process measure for which we provide no risk adjustment or stratification. We determined risk adjustment and stratification were not appropriate based on the measure evidence base and the measure construct. As a process-of-care measure, the decision to order and interpret a head CT or MRI scan within 45 minutes should not be influenced by SDS factors; rather, adjustment would potentially mask such important inequities in care delivery. Variation across patient populations is reflective of differences in the quality of care provided to the disparate patient population included in the measure's denominator.

During the measure maintenance process, we perform an annual review of the literature, to identify articles and clinical practice guidelines published in the last 12 months, which includes a scan for potential patient subpopulations for which there are differences in the clinical decision to perform a head CT or MRI scan; this most recent review identified no clear evidence of an empirical relationship between SDS and facility-level measure performance.

In addition to the evidence gathered from the literature, stakeholder feedback obtained during the three years of implementation and public reporting has not identified concerns related to SDS factors and need for risk adjustment. This supports the conceptual model upon which the measure is based.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Not applicable - No risk adjustment or stratification was performed.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable - No risk adjustment or stratification was performed.

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Not applicable - No risk adjustment or stratification was performed.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)
Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

Not applicable - No risk adjustment or stratification was performed.

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable - No risk adjustment or stratification was performed.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable - No risk adjustment or stratification was performed.

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable - No risk adjustment or stratification was performed.

2b4.9. Results of Risk Stratification Analysis:

Not applicable - No risk adjustment or stratification was performed.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable - No risk adjustment or stratification was performed.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

Not applicable - No risk adjustment or stratification was performed.

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

We tested the statistical significance of the difference between facility performance scores and the mean performance value for facilities meeting public-reporting requirements. For the 2014 data, this included 958 facilities. For each facility, the facility performance score and standard deviation was calculated. This analysis identified 43 facilities as statistical outliers. Additional details of this analysis are provided in Section **2b5.2**.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Of the 958 facilities reporting during the 2014 data collection period, 43 (4.5%) facilities had a performance value that was statistically significantly different from a mean benchmark value. Statistically meaningful difference was defined as when the facility score fell outside of the confidence interval (± 1.96 standard deviations) for the measure mean (benchmark value). Thus, this calculation identifies statistical outliers.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

Analysis of the 2014 performance data, and the subsequent rate of identification of statistically different performance for 4.5% of measured entities, demonstrates the ability of the measure to identify outlying performance. By reporting a measure mean (benchmark value), this provides an opportunity for outlying facilities to identify underperformance related to delayed interpretation of head CT or MRI and work to implement quality improvement mechanisms to increase the proportion of patients receiving rapid interpretation of head CT or MRIs scans when clinically appropriate.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications*

for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not Applicable - This measure is not risk adjusted and uses only one set of specifications.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not Applicable - This measure is not risk adjusted and uses only one set of specifications.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e.*, *what do the results mean and what are the norms for the test conducted*)

Not Applicable - This measure is not risk adjusted and uses only one set of specifications.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

This measure is calculated using chart-abstracted data. To limit the effects of missing data, abstractors cannot submit a value of "missing" for individual data elements. When they submit a value of "missing" the case is rejected from the abstraction tool. While abstractors cannot submit missing data, they may submit a value of "UTD" for select data elements for which missing information may be more likely, for example for *Time Last Known Well* before the onset of stroke symptoms. Cases where a value of "UTD" affects clinical decision making are excluded from the measure. Cases where a value of "UTD" is reflective of poor documentation are included in the denominator but excepted from the numerator. To identify the extent and distribution of cases with a value of "UTD" for a data element, we calculated the frequency of such cases as well as the distribution of cases across eligible facilities. The frequency and distribution of missing data are described in Section **2b3.3.**

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

The frequency and distribution of missing data are described in Section **2b3.3**. We did not perform statistical analyses of missing data.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

As described in Section **2b3.3** the removal of cases from the denominator and/or numerator where an abstractor submits a value of "UTD" are necessary to align with clinical guidelines and enable measure calculation. Additionally, these exclusions/exceptions limit the biasing effects of missing data. Cases where a value of "UTD" affects clinical decision making are excluded from the measure. Cases where a value of "UTD" is reflective of poor documentation are included in the denominator but excepted from the numerator. This exclusion/exception approach penalizes facilities for poor documentation, but does not artificially include cases where rapid administration of tPA may not be appropriate care.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

NQF #0661 shares key data elements with NQF #0437: Thrombolytic Therapy, which is currently an electronic clinical quality measure (eCQM). The potential for e-specification will require special attention for the Last Known Well, Date Last Known Well, Time Last Known Well, Head CT or MRI Scan Interpretation Date, and Head CT or MRI Scan Interpretation Time data elements since these currently rely on logic and inferences that abstractors have been trained to interpret. In particular, the head CT or MRI interpretation data elements, which are not part of the algorithm for NQF #0437, are not readily available in structured fields. Abstractors often rely on the radiology images or medical notes to determine the appropriate interpretation time. Additionally, electronic timestamps may not reflect the earliest interpretation time as required by the current specifications. The stroke and acute myocardial infarction expert work group (EWG) considers NQF #0661 to be wholly feasible as it is currently specified, but considers e-specification to be moderately feasible. They concur that the key data elements for NQF #0661 are not readily available in a structured format within an electronic health record (EHR). In particular, EHR systems may need a new structured field for Date Last Known Well and Time Last Known Well, which is not perceived to be a standard feature for most systems at this time.

Based on EWG feedback, EHRs will need to be compatible with RIS PACS (radiology information system and picture archiving and capture) data. If Date Last Known Well and Time Last Known Well cannot be translated into structured fields, then the data elements must be manually chart abstracted.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

No feasibility assessment **Attachment:**

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be

implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
We conducted an online survey of five members of the stroke and acute myocardial infarction (AMI) EWG with expertise in cardiology, neuro-radiology, emergency medicine, and emergency nursing to assess the face validity, feasibility, use, and usability of NQF #0661. All participants agreed or strongly agreed that patients who have a head CT or MRI scan ordered and interpreted within 45 minutes of ED arrival can be accurately identified using chart-abstracted data. Additionally, 80% of participants agreed that practical aspects of reporting this chart-abstracted measure do not place undue burden on facilities that collect the data.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
No fees, licensure, or other requirements are necessary to use this measure; however, CPT codes, descriptions, and other data are copyright 2013 American Medical Association. All rights reserved. CPT is a registered trademark of the American Medical Association. Applicable FARS\DFARS Restrictions Apply to Government Use. Fee schedules, relative value units, conversion factors, and/or related components are not assigned by the AMA, are not part of CPT, and the AMA is not recommending their use. The AMA does not directly or indirectly practice medicine or dispense medical services. The AMA assumes no liability for data contained or not contained herein.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Public Reporting Hospital Outpatient Quality Reporting (HOQR) http://www.medicare.gov/hospitalcompare/search.html Hospital Outpatient Quality Reporting https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1192804531207 Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Hospital Outpatient Quality Reporting http://www.medicare.gov/hospitalcompare/search.html Hospital Outpatient Quality Reporting https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1192804531207

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Public Reporting:
Name of program and sponsor: The CMS HOQR Program
Purpose: The HOQR Program is a pay for quality data reporting program implemented by CMS for outpatient hospital services. In addition to providing hospitals with a financial incentive to report their quality of care measure data, the HOQR Program provides CMS with data to help Medicare beneficiaries make more informed decisions about their health care. Hospital quality of care information gathered through the HOQR Program is publicly available on the Hospital Compare website.
Accountable entities and patients: The publicly reported values (on Hospital Compare) are calculated for all facilities in the United States that meet minimum case count requirements. During the October 2012 through September 2013 data collection period, 918 facilities met the minimum case count. During the October 2013 through September 2014 data collection period, 959 facilities met the minimum case count. Facilities eligible to report this measure are subject to the Outpatient Prospective Payment System (OPPS) guidelines.

Quality Improvement with Benchmarking (external benchmarking to multiple organizations):
Name of program and sponsor: The CMS HOQR Program

Purpose: The HOQR Program is a pay for quality data reporting program implemented by CMS for outpatient hospital services. In addition to providing hospitals with a financial incentive to report their quality of care measure data, the data is publicly reported on the Hospital Compare Website. The data reported on Hospital Compare not only shows the hospital's score on the measure, but also provides state and national averages for the measure. This enables consumers to compare the hospital's performance to other facilities and determine if the facility is an outlier.
Accountable entities and patients: The publicly reported values (on Hospital Compare) are calculated for all facilities in the United States that meet minimum case count requirements. During the October 2012 through September 2013 data collection periods, 918 facilities met the minimum case count. During the October 2013 through September 2014 data collection period, 959 facilities met the minimum case count. Facilities eligible to report this measure are subject to the OPPS guidelines.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
This measure is publicly reported.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
This measure is publicly reported.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Summary statistics of performance scores during the October 2012 through September 2013 and October 2013 through September 2014 data collection periods are provided in Section 1b.2.

The median rate of head CT or MRI scans interpreted within 45 minutes of ED arrival, given that the patient had a principal diagnosis associated with acute ischemic or hemorrhagic stroke, arrived at the ED within two hours of the time last known well, and had an order for a head CT or MRI scan has increased 14.5% (62.0% to 71.0%) between 2012 and 2014. Nine hundred eighteen facilities met minimum case counts during the October 2012 through September 2013 data collection periods, and 959 facilities met minimum case counts during the October 2013 through September 2014 data collection period. During the October 2012 through September 2013 data collection period, there were 16,817 sampled cases where a patient had a principal diagnosis associated with acute ischemic or hemorrhagic stroke, who arrived at the ED within two hours of the time last known well, and who had an order for a head CT or MRI scan. Of those patients, 10,026 had a head CT or MRI scan interpreted within 45 minutes of ED arrival (59.6%). During the October 2013 through September 2014 data collection period, 17,108 patients had a principal diagnosis associated with acute ischemic or hemorrhagic stroke, who arrived at the ED within two hours of the time last known well, and who had an order for a head CT or MRI scan. Of those patients, 11,534 had a head CT or MRI scan interpreted within 45 minutes of ED arrival (67.4%).
These cases reflect only a subset of the patients eligible for the measure. Dependent upon the facility's total case

count, the facility may report all cases or a sample of cases; thus, the number of patients receiving high-quality healthcare as performance on the measure improves is larger than the number of cases captured by the measure.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not applicable, as there is demonstrated improvement in measure performance over time.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We did not identify any unintended consequences during measure testing. Similarly, no evidence of unintended consequences to individuals or populations has been reported by external stakeholders since its implementation. We will continue to monitor the potential for unintended consequences through an annual review of the literature as well as an ongoing review of stakeholder comments and inquiries.The risk in advancing measures that address timeliness is that there may be a decrease in testing performance to avoid measurement, however this is not likely due to the need to assess diagnostic results to ensure a proper diagnosis.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0437 : STK 04: Thrombolytic Therapy

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
Diagnosis and treatment of ischemic stroke: percentage of patients with stroke symptoms who undergo a CT scan within 25 minutes of arrival in the emergency department - Institute for Clinical Systems Improvement (ICSI)

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Although NQF #0437 (used in the Hospital Inpatient Quality Reporting [HIQR] Program) is similar to NQF #0661 (HOQR), the two measures serve different target populations and purposes: the HOQR measure focuses on imaging in the ED setting, while the HIQR measure focuses on administration of thrombolytic therapy in an inpatient setting. Both measures do, however, share a number of key data elements (i.e., Last Known Well, Date Last Known Well, Time Last Known Well, and Arrival Time). The specifications for the two measures are generally aligned, where possible. As appropriate, the measure maintenance team for the HOQR measure (NQF #0661) incorporates specification updates added by the measure maintenance team for the HIQR measure (NQF #0437) to maintain harmonization (e.g., updates to the appropriate ICD-10 codes to determine measure inclusion). The measure-maintenance teams for both reporting programs meet periodically to resolve any inconsistencies in the interpretation or guidance provided for the shared data elements.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
We did not identify any competing measures that address both the same measure focus and target population as NQF #0661.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
No appendix **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services
Co.2 **Point of Contact:** Megan, Hayden, megan.hayden@cms.hhs.gov, 410-786-1970-
Co.3 **Measure Developer if different from Measure Steward:** The Lewin Group
Co.4 **Point of Contact:** Colleen, McKiernan, Colleen.McKiernan@lewin.com, 703-269-5595-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The contractor has convened an EWG, which evaluates and provides feedback on measure-development and maintenance efforts for a set of six stroke and AMI measures. Specifically, the EWG provides direction and feedback through all phases of project activities, including expansion of the measures to additional CMS quality reporting programs, updates to the current specifications of these six measures, review of quantitative testing results, feedback on qualitative testing questions (i.e., results of EWG member questionnaires), and support for

endorsement of the measures by the National Quality Forum (NQF).

The following is a list of the contractor's EWG members:

Joseph P. Drozda, Jr., MD
TEP 2010; Mercy Health, Rep. of American College of Cardiology; Director of Outcomes Research

Mustapha Ezzeddine, MD
University of Minnesota Medical Center, Director, Stroke Program

T. Bruce Ferguson, Jr., MD, FACC
TEP 2010; Brody School of Medicine at ECU, Dept. of Cardiovascular Sciences, Professor of Surgery and Physiology

Joseph V. Messer, MD, MACC
TEP 2010; Rush University Medical Center, Rep. of American Medical Association, Professor of Medicine

Cathy Olson, MSN, RN
Emergency Nurses Association (ENA), Institute for Quality, Safety, and Injury Prevention, Director

David Seidenwurm, MD
American Society of Neuroradiology (ASNR); American College of Radiologists (ACR)

Stephen Traub, MD
TEP 2010; Mayo Clinic, Department of Emergency Medicine, Chair

Paul D. Varosy, MD, FACC, FAHA, FHRS
TEP 2010; VA Eastern Colorado Health Care System, Director of Cardiac Electrophysiology

Matt Zavadsky, MS-HPA
National Association of Emergency Medical Technicians (NAEMT)

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2012
Ad.3 **Month and Year of most recent revision:** 01, 2016
Ad.4 **What is your frequency for review/update of this measure?** Annually
Ad.5 **When is the next scheduled review/update for this measure?** 01, 2017

Ad.6 **Copyright statement:** This measure does not have a copyright.
Ad.7 **Disclaimers:** CPT codes, descriptions, and other data only are copyright 2013 American Medical Association. All rights reserved. CPT is a registered trademark of the American Medical Association. Applicable FARS\DFARS Restrictions Apply to Government Use. Fee schedules, relative value units, conversion factors and/or related components are not assigned by the AMA, are not part of CPT, and the AMA is not recommending their use. The AMA does not directly or indirectly practice medicine or dispense medical services. The AMA assumes no liability for data contained or not contained herein.

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 1814
Measure Title: Counseling for Women of Childbearing Potential with Epilepsy
Measure Steward: American Academy of Neurology
Brief Description of Measure: All female patients of childbearing potential (12–44 years old) diagnosed with epilepsy who were counseled or referred for counseling for how epilepsy and its treatment may affect contraception OR pregnancy at least once a year
Developer Rationale: From original NQF submission -
Educate women about epilepsy and how its treatments may affect contraception and pregnancy. This will inform women of childbearing potential about the risks of epilepsy and AED therapy prior to pregnancy. It will provide an opportunity to educate this population about folic acid supplementation, monotherapy, medication alternatives and how to obtain obstetrical, prenatal and pregnancy care. This measure will help them understand the risk and mitigate the risks which may prevent fetal malformation, unplanned pregnancies and improve the patients´ quality of life.

Numerator Statement: Female patients or caregivers counseled* at least once a year about how epilepsy and its treatment may affect contraception OR pregnancy.

*Counseling should include a discussion about folic acid supplementation, contraception, potential anti-seizure medications effect(s) on pregnancy, safe pregnancies, and breastfeeding.
Denominator Statement: All females of childbearing potential (12-44 years old) with a diagnosis of epilepsy.
Denominator Exclusions: Excluded: patients diagnosed with menopause or surgically sterile.

Exceptions:
Patient has a diagnosis of neurodevelopmental disorder, encephalopathy, hydrocephalus, brain injury, or cerebral palsy.

Patient has a diagnosis of severe cognitive impairment or severe intellectual disability.

Measure Type: Process
Data Source: Electronic Clinical Data : Electronic Health Record, Paper Medical Records
Level of Analysis: Clinician : Group/Practice

IF Endorsement Maintenance – Original Endorsement Date: Mar 06, 2013 **Most Recent Endorsement Date:** Mar 06, 2013

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion
.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence..

- The developer states that eligible patients whom are counseled on how epilepsy and its treatment and how this may affect contraception or pregnancy >would increase folic acid use and receive appropriate obstetrical , prenatal and pre-pregnancy care in accordance with their wishes > would decrease the number of unintended pregnancies, complications during pregnancies and malformations in infants and increase number of children born at term.

Systematic Review of the evidence specific to this measure?	☒ **Yes**	☐ **No**
Quality, Quantity and Consistency of evidence provided?	☒ **Yes**	☐ **No**
Evidence graded?	☒ **Yes**	☐ **No**

Evidence Summary and **Summary of prior review in 2012**

The evidence provided as part of the previous evaluation included:

- QUIET Indicator Study (Pugh et al. 2011) on quality of care for adults with epilepsy highlights the gap in care for women's issues with very low concordance rates. Overall of the 111 women of childbearing potential with 128 opportunities for recommended care, only 36.72% of the time quality indicator concordant care was provided.
- Women should be given information about contraception, conception, pregnancy and breastfeeding. Information should be given in advance of sexual activity or pregnancy. (Level C)(NICE 2004)
- Couldridge and colleagues reviewed evidence (including non-RCT studies) on the information and counseling needs of people with epilepsy, including the preferred format, timing, and delivery of information and counseling, and the outcomes of information giving and counselling.
- Crawford and Lee (1999) surveyed women and found that overall, women felt there was a need for more information about epilepsy and pregnancy. The survey concluded that women with epilepsy (WWE) wanted, and needed, more information and counseling about issues relating to contraception, pregnancy, and the menopause.
- Crawford and Hudson (2003) published the findings of the Ideal World survey that identified the most important issues for WWE aged 19 to 44 years who were considering having children. These include risk of epilepsy/medication affecting the unborn child, effect of pregnancy on seizure control and risk of a child developing epilepsy.
- Other prior guidelines that are graded can be found in the A-F listing.

- In the 2012 NQF endorsement evaluation, the Committee disagreed about the level of evidence. Several members noted that the submission did not provide evidence of a direct link between counseling and patient outcomes. Others noted that the submission did provide evidence that epilepsy treatment can affect both contraception and child development during pregnancy, as well as evidence that women with epilepsy feel that they are not getting adequate information about pregnancy
- The Committee recommended invoking the evidence exception because the measure impacts a specific population and there is the potential for great harm if such counseling is not done. Members almost unanimously agreed to invoke the exception to the evidence sub-criterion.
- In the last submission, the developer stated that the guideline/indicator authors did not provide an explicit process or documentation of a process whereby precision and directness were detailed in a systematic review to demonstrate the quality of the body of evidence for this measure.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates:

- NICE has updated guidelines (2014) that are rated at a ***Level III*** (well-designed non-experimental descriptive studies, case-control studies, and case studies). These include guidelines on counseling about contraception, conception, pregnancy, caring for children and breastfeeding, and menopause; information that should also be

given to people who are closely involved with women and girls with epilepsy; information on contraception, and role of healthcare professionals knowledge in becoming familiar with relevant information and the availability of counseling.

- Although the updated guidelines provided relevant information, there is no evidence that links counseling (with this information) of WWE and positive clinical outcomes.

Exception to evidence --N/A

Guidance from the Evidence Algorithm

Process measure with evidence from graded systematic reviews (Box 3) → QQC provided for Harden, et al. recommendations but not NICE recommendations (Box 4) → Harden, et al has low consistency (Box 5c); NICE not graded high quality (Box 6) → LOW

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

Preliminary rating for evidence: ☐ High ☐ Moderate ☒ Low ☐ Insufficient

Rationale: Because QQC not provided for NICE guideline recommendations, rating is primarily based on the Level III grade of the NICE evidence. Additional information on NICE QQC could elevate rating to MODERATE.

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- AAN (2013) tested its Women with Epilepsy of Childbearing potential measure and evidence of a gap in care remains. Data from the testing project showed that on average less than 40% of women received counseling about epilepsy and how its treatment may affect contraception and pregnancy.
- The QUality Indicators for Epilepsy Treatment in adults (QUIET) study demonstrated that only 34% of female patients receive counselling on aspects of epilepsy care specific to women (neurologist alone=32.88%; shared (neurologists and primary care=44.83%; and primary care alone=11.11%).
- Knowledge about the use of seizure medications during pregnancy was low, with less than half of neurologists able to identify which medications were linked to adverse events during pregnancy.

Disparities

- In the previous submission, the developer cited literature indicating racial disparities in the prevalence and treatment of epilepsy but no information about disparities in counseling about pregnancy. No updated information on disparities for the measure as specified was provided.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of disparities in counseling women with epilepsy that you are aware of?*

Preliminary rating for opportunity for improvement: ☒ High ☐ Moderate ☐ Low ☐ Insufficient

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **There is no evidence supporting that counseling will improve outcomes in terms of preventing unwanted pregnancy or preventing fetal or maternal complications for WWE.

**When the measure was first proposed in 2012, there was little evidence that counseling leads to better outcomes. Four years

later, there still remains no evidence.

1b. Performance Gap

Comments: **Evidence support that women want more counseling and information.

**Performance data is provided from three neurology practices in Minnesota. Only 37% of patients received counseling on both contraception and pregnancy.

1c. High Priority (previously referred to as High Impact)

Comments: **

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- Data source: EHRs, Paper Medical Records. Although not specified, data appears to be available via registry as well.
- Numerator: Female patients or caregivers counseled* at least once a year about how epilepsy and its treatment may affect contraception OR pregnancy.*Counseling should include a discussion about folic acid supplementation, contraception, potential anti-seizure medications effect(s) on pregnancy, safe pregnancies, and breastfeeding.
- Denominator: All females of childbearing potential (12-44 years old) with a diagnosis of epilepsy
- Denominator Exclusion: patients diagnosed with menopause or surgically sterile.
- Exceptions: Patient has a diagnosis of neurodevelopmental disorder, encephalopathy, hydrocephalus, brain injury, or cerebral palsy. Patient has a diagnosis of severe cognitive impairment or severe intellectual disability.
- Level of analysis is at the clinician group or practice level
- ICD-9 and updated ICD-10 codes are provided.

2a2. Reliability Testing Testing attachment

Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- At the time of the prior evaluation, this measure had not been tested for reliability or validity. The measure was accepted for consideration for potential time-limited endorsement because it is a process measure, was already included in a CMS quality improvement program (PQRS), and because it fills a gap area in NQF's measure portfolio. Before accepting this measure for evaluation during the prior project, NQF asked developers to confirm their intent to complete reliability and validity testing within 12 months if granted time-limited endorsement. When voting on the Scientific Acceptability of this untested measure, the Committee considered whether the measures were precisely specified and whether the measure specifications are consistent with the evidence presented for the measure (voting either "yes" or "no").

Updated testing:

- The developer states that they identified and recruited three Neurology practices in Minnesota that have experience treating patients with epilepsy. Three Neurology practices volunteered to participate and submit retrospective data from the 2012 calendar year (i.e. dates of service 01/01/2012 – 12/31/2012). One practice extracted all data via EHR; two practices had data extracted via EHR and also required manual chart review.
- The developer described a 3-step process that was used to demonstrate the accuracy of the data abstracted from the three medical group practices.
- NOTE: This was the testing data submitted to CSAC to support change from time-limited to full endorsement.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☐ **No**

Method(s) of reliability testing

.

- The developer states that a validation of the data is done through: 1) denominator certification, 2) data file quality checks, and 3) validation audit. A comparison between the medical group practices and this 3-step process including the validation audit is considered the gold standard. This type of testing is considered by NQF to be *validity testing* of the data elements; however, it will satisfy NQF requirements for data element reliability testing.

Results of reliability testing

See results of data element validity testing below.

Guidance from the Reliability Algorithm:

Specifications precise (Box 1) > No empirical reliability testing (Box 2) > Empirical validity testing conducted (Box 3) > Data-element validity testing (Validity algorithm, Box 10) > comparison against gold standard, but only percent agreement statistics provided (Validity algorithm, Box 11) > recommend low

Questions for the Committee:

- *Based on the process describe are not confident that the data is reliability using patient level data?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☐ **Moderate** ☒ **Low** ☐ **Insufficient**

Rationale: Only percent agreement statistics were provided for site B (other statistics for sites A and C not needed, given 100 percent agreement). Additional statistics for site B (e.g., kappa statistics) could elevate rating to MODERATE.

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

As noted above, there was no validity testing of the measure at the time of the 2012 evaluation

Describe any updates to validity testing:

- Validity testing involved validation of data that was abstracted by three Minnesota neurology practices. These

practices volunteered to participate and submit retrospective data from the 2012 calendar year (i.e. dates of service 01/01/2012 – 12/31/2012). This was the testing data submitted to CSAC to support change from time-limited to full endorsement.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Validity testing method:

- The developer recruited three neurology practices in Minnesota that have experience treating patients with epilepsy. As part of the process, a data collection guide, measure flow, and detailed file specifications to educate and assist each medical group in the data collection and submission process was provided.
- As a requirement of participating in the measure testing each group had to submit denominator certification forms. This process helps ensure that each medical group is using the appropriate measure parameters and collecting data in a standardized way.
- One neurology practice extracted all data via EHR; and two practices had data extracted via EHR and also required manual chart review.
- Developer states that MNCM (contractor) completed validation of the data in a three-step process: 1) denominator certification, 2) data file quality checks, and 3) validation audit.
    - In the first step of denominator certification, each participant needed to attest to submitting data of the denominator that follows measure specifications. This included diagnosis codes, etc.
    - In the data file quality checks, the contractor reviewed the data field by row for each patient . Details of the specifics of this step.
    - Finally, audits were completed of the patient records to verify the submitted clinical data. A validation process developed by the NCQA, known as the "8 and 30" was used. In this process, the first eight records are verified for accuracy and if no errors are identified, the data is considered to be 100% compliant.

The testing sample Included:

Neurology Groups	A	B	C	Total
No. of Patients	751	581	59	1391
No. of Providers /Clinicians	9	36	22	67
Records reviewed during audit	11	30	8	49

Validity testing results:

- There were no major flaws or issues identified issues during the review of each medical group's denominator forms and each medical group passed denominator certification within the given timeframe. A few issues needing follow-up were identified.
- The developer states that the issues identified through the data file quality checks were generally minor, requiring no corrections. See comments.
- The clinical data audit revealed some data errors, requiring one medical group to make corrections and resubmit data. Individual medical group results were as follows. The results are detailed here.
    - For medical group A, 8 records were reviewed (100% compliant), and 3 additional records were reviewed for "other reason patient not counseled" (66%compliant). One record indicated that patient had functional seizures and not epilepsy but should have been counted as no counseling provided.
    - For medical group B, 30 records were reviewed and 26 were compliant (87%). Three records had a code "2" for no counseling due to intellectually disability (ID) however the diagnosis could not be verified in

the record and "1" record reported as counseling could not be verified. In this medical group, they verified that patients who were submitted were either as surgically sterile or who had an intellectual disability and were resubmitted with corrections.

- In group C, 8 records were reviewed, all 8 were compliant. One record could have been flagged for a medical reason to not receive counseling but because these reasons were to either surgically sterile or ID, it was appropriate that these could have been counseled and were not counted as errors.

Developer states that the process is successful in identifying errors (with subsequent corrections) and verifying the accuracy of the data submitted by medical groups A, B, and C and states that there is no significant flaws or errors with the data.

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Other specific question of the validity testing?*
- *Do you think the method as explained demonstrates adequate validity of the data elements included in the measure?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

Exclusion include:

- Patient was surgically sterile (tubal ligation, hysterectomy)
- Patient has an intellectual disability as defined by ICD-9 codes
- For these two exclusion criteria, 196 or 14.1% (196/1391) patients were excluded from the denominator.
- The developer also discussed those patients for whom counseling was not provided due to "other medical reason". This was indicated by a code and accompanying description. According to the developer, these reasons were not used to exclude patients from the measure; rather the purpose was to provide additional information about the population of patients included in the measure. Reasons Provided by Medical Groups for Not Providing Counseling include:

- The contractor raised concerns regarding the denominator and intent of the measure. They suggest that possibly adding a component indicating that the patient is sexually active or has the potential to be sexually active, and not physically handicapped. Rather than trying to capture/ code every possible exclusion; this may be an option.

_____

Questions for the Committee:

- *Do you have any concerns regarding the testing of exclusions?*
- *How are patients qualifying for exceptions (e.g., neurodevelopmental disorder, encephalopathy, hydrocephalus, brain injury, or cerebral palsy) handled in the measure. Do you agree?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*
- *Do your think changing exclusions as discussed by the contractor would improve the measure?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- No performance data for this measure as specified was provided. CMS Physician Quality Reporting System (PQRS) does report that 57.5% of eligible neurologists participated in PQRS reporting in 2013. Data on this measure not known, as it is not reported in the PQRS and ERX Experience Report.
- The American Academy of Neurology Axon Registry has an internal benchmarking quality registry with plans to expand to external benchmarking. This tool enables neurology practices to identify and improve gaps in the quality of neurologic care. The Axon Registry was launched in Q3 of 2015. There are plans for this measure to be incorporated into the registry in 2016.

- Currently there are 3 pilot cohorts in progress providing data to the registry. Axon Registry is currently being piloted and will be available to all AAN US members. Currently, there are 42 practices (487 neurology providers) in the 3 pilot cohorts.

2b6. Comparability of data sources/methods:

Developer did not provide.

2b7. Missing Data

- The developer states that if data is missing from denominator, the case is deleted. If data for denominator is met then case information is included for the measure.
- The developer also cited that validation process by MNCM.

- **Guidance from Algorithm:** Measure specification consistent (Box 1)> Some potential threats analyzed (Box 2)>Empirical Validity testing conducted (Box 3)>validity not conducted with performance measure score (Box 6)>Validity tested at patient level data elements (Box 10)>Only percent agreement provided (Box 11)> recommend INSUFFICIENT

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

Rationale: Only percent agreement statistics were provided for site B (other statistics for sites A and C not needed, given 100 percent agreement). Additional statistics for site B (e.g., kappa statistics) could elevate rating to MODERATE, depending on perception of this site having to resubmit data. Additional information about meaningful differences and missing data would be useful.

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **How women are counseled is not addressed so whether Women get the type of counseling and information they desire is not addressed. What is address is simply were they counseled.

2a2. Reliability Testing

Comments: **Whether the exclusions are valid based on the data provided appears to be a problem and the best possible ways to not incorrectly assume a woman does not need counselling on contraception or pregnancy.

**Validity testing revealed errors, requiring one of the medical practices to resubmit data. The contractor performing the validity testing had concerns about the denominator specifications and the intent of the measure.

2b2. Validity Testing

Comments: **

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Reliability testing was probably not sufficient due to failure to more statistically evaluate cases that were considered to be incorrect in terms of exclusion

**Reliability testing was performed at the data element level, via validity testing.

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Developer stated that the data is generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, and coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims).
- Abstracted from a record by someone other than person obtaining original information for quality measure or registry.
- All data elements are in defined fields in a combination of electronic sources
- Data are being successfully pulled from EHRs without need for manual extraction.
- According to developer, this information will be used to strengthen the measure in the next scheduled update in Q4 2016.
- The AAN´s Axon Registry was initiated in 2015 Q3.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **It appears that in the clinics studied much of the information was presented but required review of the charts manually. I wonder how many providers use fields to chart vs. their own templates which may make it difficult to generalize
**The data elements require manual abstraction.

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

The developer describes three reporting programs in place:

- CMS Physician Quality Reporting System (PQRS) 57.5% of eligible neurologists participated in PQRS reporting in 2013. Data on this measure not known, as not reported in the PQRS and ERX Experience Report.
- The American Academy of Neurology Axon Registry has an internal benchmarking quality registry with plans to expand to external benchmarking. This tool enables neurology practices to identify and improve gaps in the

quality of neurologic care. Axon Registry was launched in Q3 of 2015. Plans for this measure to be incorporated into the registry in 2016.

- Currently there are 3 pilot cohorts in progress providing data to the registry. Axon Registry is currently being piloted and will be available to all AAN US members. Currently, there are 42 practices (487 neurology providers) in the 3 pilot cohorts.

Improvement results

- Currently there is no data on improvement.

Unexpected findings (positive or negative) during implementation

No finding reported

Potential harms

No harm mentioned

Feedback :

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4. Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **Currently being tested on the AAN Axon registry which is to allow groups to benchmark their performance and ultimatly provide between group comparisons.

Used for CMS´s Physician Quality Reporting System (PQRS) but not the granularity to look at this measure.

**The measure is not publicly reported. It is available to be used as part of PQRS, but it is not clear to what extent the measure is actually being used.

Criterion 5: Related and Competing Measures

Related or competing measures

There are no completing or related measures.

Harmonization

N/A

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1814
Measure Title: Counseling for Women of Childbearing Potential with Epilepsy
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 1/15/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome

☐ Health outcome: Click here to name the health outcome
☐Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Counseling for Women of Childbearing Potential with Epilepsy

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

Not Applicable

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Not Applicable

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

Process	Intermediate Outcome	Outcome
Eligible patients counseled on how epilepsy and its treatment may affect contraception or pregnancy	Increased folic acid use in eligible population Eligible patients receive appropriate obstetrical, prenatal and pre-pregnancy care in accordance with their wishes	Decreased number of unintended pregnancies for eligible population Decreased complications during pregnancy for eligible patients Increased number of children born at term for eligible patients Decreased malformations in infants born to eligible patients

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

1. Pugh MJ, Berlowitz DR, Montouris G, et al. What constitutes high quality of care for adults with epilepsy? Neurology 2007;69:2020-2027

2. Harden CL, Hopp J, Ting TY, et al. Practice Parameter update: Management issues for women with epilepsy-Focus on pregnancy (an evidence-based review): Obstetrical complications and change in seizure frequency: Report of the Quality Standards Subcommittee and Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology and American Epilepsy Society. Neurology 2009;73:126-132.
3. Harden CL, Meador KJ, Pennell PB, et al. Practice Parameter update: Management issues for women with epilepsy – Focus on pregnancy (an evidence-based review): Teratogenesis and perinatal outcomes: Report of the Quality Standards Subcommittee and Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology and American Epilepsy Society. Neurology 2009;73:133-141.
4. National Institute of Clinical Health and Excellence. The epilepsies: the diagnosis and management of the epilepsies in adults and children in primary and secondary care (update). 2012. Clinical guideline 137. Available at: http://www.nice.org.uk/Guidance/cg137 Accessed on February 18, 2014.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Since the release of the original measure, the guideline statements have been updated, and additional guideline statements added supporting the need for counseling.

A. If a woman with epilepsy is of childbearing potential and receives oral contraceptives in conjunction with an enzyme inducing AED [Antiepileptic Drug], THEN decreased effectiveness of oral contraception should be addressed. (higher doses of the oral contraceptive, alternative birth control methods, or change AED). (Level A 2++/Primary)1
B. Patients with epilepsy should receive an annual review of information including topics such as: … Contraception, family planning, and how pregnancy and menopause may affect seizures (evidence grade C)1
C. Women with epilepsy (WWE) should be counseled that seizure freedom for at least 9 months prior to pregnancy is probably associated with a high rate (84%-92%) of remaining seizure-free during pregnancy. (Level B) 2
D. Women with epilepsy who smoke should be counseled that they possibly have a substantially increased risk of premature contractions and premature labor and delivery during pregnancy. There is possibly a substantially increased risk of premature contractions and premature labor and delivery during pregnancy for WWE who smoke. (Level C)2
E. Counseling of WWE who are contemplating pregnancy should reflect that there is probably no increased risk of reduced cognition in the offspring of WWE not taking AEDs (Level B).3
F. To reduce the risk of MCMs, avoidance of the use of VPA during the first trimester of pregnancy, if possible, may be considered, compared to the use of PHT or LTG. [MCMs=major congenital malformations; VPA=valproate; PHT=phenytoin; LTG=lamotrigine] (Level C)3
G. In order to enable informed decisions and choice, and to reduce misunderstandings, women and girls with epilepsy and their partners, as appropriate, must be given accurate information and counselling about contraception, conception, pregnancy, caring for children and breastfeeding, and menopause. (Level III)4
H. Information about contraception, conception, pregnancy, or menopause should be given to women and girls in advance of sexual activity, pregnancy or menopause, and the information should be tailored to their individual needs. This information should also be given, as needed, to people who are closely involved with women and girls with epilepsy. These may include her family and/or carers. (Level III)4
I. All healthcare professionals who treat, care for, or support women and girls with epilepsy should be familiar with relevant information and the availability of counselling. (Level III)4
J. Discuss with women and girls of childbearing potential (including young girls who are likely to need treatment into their childbearing years), and their parents and/or carers if appropriate, the risk of AEDs causing malformations and possible neurodevelopmental impairments in an unborn child. Assess the risks and benefits of treatment with individual drugs. There are limited data on risks to the unborn child associated with newer drugs. Specifically discuss the risk of continued use of sodium valproate to the unborn child, being aware that higher doses of sodium valproate (more than 800 mg/day) and polytherapy, particularly with sodium valproate, are associated with greater risk. (Evidence comes from three systematic reviews; one review focused on incidence of malformation and the other two on child neurodevelopmental outcomes. No individual RCTs were reviewed. This recommendation was also based on GDG consensus opinion.)4

K. In women of childbearing potential, the possibility of interaction with oral contraceptives should be discussed and an assessment made as to the risks and benefits of treatment with individual drugs. (Level III)4
L. In girls of childbearing potential, including young girls who are likely to need treatment into their childbearing years, the possibility of interaction with oral contraceptives should be discussed with the child and/or her carer, and an assessment made as to the risks and benefits of treatment with individual drugs. (Level III)4
M. In women and girls of childbearing potential, the risks and benefits of different contraceptive methods, including hormone-releasing intrauterine devices (IUDs), should be discussed. (Level III)4
N. If a woman or girl taking enzyme-inducing AEDs chooses to take the combined oral contraceptive pill, guidance about dosage should be sought from the SPC and current edition of the BNF (available at http://bnf.org External Web Site Policy). (Level III)4
O. Women and girls with epilepsy need accurate information during pregnancy, and the possibility of status epilepticus and sudden death in epilepsy (SUDEP) should be discussed with all women and girls who plan to stop AED therapy (see the section 'Withdrawal of Pharmacologic Treatment' above).4

1a.4.3. Grade assigned to the quoted recommendation <u>with definition</u> of the grade:

A. Level A 2++/Primary A: Rated as appropriate
B. Evidence grade C
C. Level B = Probably effective, ineffective or harmful (or probably useful/predictive or not useful/predictive) for the given condition in the specified population. (Level B rating requires at least one Class I study or two consistent Class II studies.
D. Level C = Possibly effective, ineffective or harmful (or possibly useful/predictive or not useful/predictive) for the given condition in the specified population. (Level C rating requires at least one Class II study or two consistent Class III studies
E. Level B = Probably effective, ineffective or harmful (or probably useful/predictive or not useful/predictive) for the given condition in the specified population. (Level B rating requires at least one Class I study or two consistent Class II studies.
F. Level C = Possibly effective, ineffective or harmful (or possibly useful/predictive or not useful/predictive) for the given condition in the specified population. (Level C rating requires at least one Class II study or two consistent Class III studies.
G. Level III-Well-designed non-experimental descriptive studies, case-control studies, and case studies
H. Level III-Well-designed non-experimental descriptive studies, case-control studies, and case studies
I. Level III-Well-designed non-experimental descriptive studies, case-control studies, and case studies
J. Following statement provided verbatim, “Evidence comes from three systematic reviews; one review focused on incidence of malformation and the other two on child neurodevelopmental outcomes. No individual RCTs were reviewed. This recommendation was also based on GDG consensus opinion.”
K. Level III-Well-designed non-experimental descriptive studies, case-control studies, and case studies
L. Level III-Well-designed non-experimental descriptive studies, case-control studies, and case studies
M. Level III-Well-designed non-experimental descriptive studies, case-control studies, and case studies
N. Level III-Well-designed non-experimental descriptive studies, case-control studies, and case studies
O. Women and girls with epilepsy need accurate information during pregnancy, and the possibility of status epilepticus and sudden death in epilepsy (SUDEP) should be discussed with all women and girls who plan to stop AED therapy (see the section 'Withdrawal of Pharmacologic Treatment' above).4

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

Pugh -

Ratings

1-3 clearly appropriate/ reliable/ necessary

4-6 uncertain or equivocal

7-10 appropriate/ reliable/ necessary

Levels of Evidence: [31]

1++ High quality meta-analyses, systematic reviews of RCTs or RCTs with low risk of bias

1+ Well conducted meta analyses, systematic reviews, or RCTs with low risk of bias

2++ High quality systematic reviews or studies without randomization, one or more high quality case-control or cohort study with low risk of confounding and high probability of a causal relationship

2+ Well conducted case-control or cohort study with low risk of confounding and moderate probability that the relationship is causal

3 Non-analytic studies (case reports/ case series)

4 Expert opinion

AAN Classification Recommendations
Appendix e-5: Classification of Recommendations

A = Established as effective, ineffective or harmful (or established as useful/predictive or not useful/predictive) for the given condition in the specified population. (Level A rating requires at least two consistent Class I studies.)*

B = Probably effective, ineffective or harmful (or probably useful/predictive or not useful/predictive) for the given condition in the specified population. (Level B rating requires at least one Class I study or two consistent Class II studies.)

C = Possibly effective, ineffective or harmful (or possibly useful/predictive or not useful/predictive) for the given condition in the specified population. (Level C rating requires at least one Class II study or two consistent Class III studies.)

U = Data inadequate or conflicting; given current knowledge, treatment (test, predictor) is unproven.

*In exceptional cases, one convincing Class I study may suffice for an "A" recommendation if 1) all criteria are met, 2) the magnitude of effect is large (relative rate improved outcome > 5 and the lower limit of the confidence interval is > 2).

NICE Evidence

NICE National Collaborating Centre for Primary Care. The diagnosis and management of the epilepsies in adults and children in primary and secondary care. London (UK): Royal College of General Practitioners; 2004 Oct.

Rating Scheme for Strength of the Evidence
Ia-Systematic review or meta-analysis of randomized controlled trials
Ib-At least one randomized controlled trial
IIa-At least one well-designed controlled stud without randomization
IIb-At least one well-designed quasi-experimental descriptive studies, such as a cohort study
III-Well-designed non-experimental descriptive studies, case-control studies, and case studies

IV-Expert committee reports, opinions and/or clinical experience of respected authorities

Rating Recommendations

A* Directly based on category I evidence (meta-analysis of randomized controlled trials (RCTs) or at least one RCT)
B* Directly based on category II evidence (at least one controlled study without randomization or at least one other quasi-experimental study) or extrapolated from category I evidence
C* Directly based on category III evidence (non-experimental descriptive studies) or extrapolated from category I or II evidence
D* Directly based on category III evidence (expert committee reports or opinions and/or clinical experience of respected authorities) or extrapolated from category I, II or III evidence
N Recommendation taken from NICE guideline or technology appraisal guidance

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**
Pugh Same as above
AAN URL: http://www.neurology.org/content/suppl/2009/04/27/WNL.0b013e3181a6b325.DC1.html
NICE Same as above

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☒ Yes → ***complete section 1a.7***
☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):
Not Applicable

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.
Not Applicable

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:
Not Applicable

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)
Not Applicable

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**
Not Applicable

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):
Not Applicable

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**
Not Applicable

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of

evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

All guideline groups evaluated various forms of the question, do women with epilepsy have an increased risk of pregnancy-related complications? As well as an evaluation of evidence on seizure recurrence and birth defect potential.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

See above evidence grades and verbatim recommendations

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

See 1a.4.4

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

Pugh data previously provided in prior submissions.

AAN guidelines range - 1985-February 2008

NICE range June 2010 to September 2013 (Builds off of past NICE searches of evidence)

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Details on Pugh data previously provided.

AAN Obstetrics guideline identified 9 studies graded Class III or higher. Regarding premature contractions and premature labor and delivery: "One Class I study5 showed no substantially increased risk of premature contractions or premature labor and delivery in WWE taking AEDs compared to control women without epilepsy (OR 0.51, 95% CI 0.19 – 1.36). One Class II study12 showed an increased risk for WWE who were smokers compared to control women who were also smokers (OR 3.4, 95% CI
1.8–6.5) (data not given for all WWE compared to controls). One Class III study13 also showed an increased risk (*p* _ 0.05). Another Class III study8 demonstrated no significant increased risk but was insufficiently sensitive to exclude a substantially increased risk (OR 8.24, 95% CI 0.92–70.32). A Class III study11 showed no significant increased risk but was not sufficiently sensitive to exclude an increased risk (RR 0.7, 95% CI 0.3–1.4). In a categorical,_2 statistic, it was reported that the rates of premature births were not different than controls (*p* _ 0.3),9 and another study found no differences in gestational ages in the offspring of WWE compared to controls (WWE _ 38.06, SD 1.42 vs controls _ 38.17, SD 3.58 weeks)."

Regarding Seizure Recurrence in previously seizure free WWE: "Two Class II articles16,17 showed that for WWE who were seizure-free for 9 months prior to pregnancy, 84%–92% remained seizure-free during pregnancy (table e-4). In one study, 38 of 45 (84%; CI 0.71– 0.92) pregnant WWE remained seizure-free,16 and in the other study, 47 of 51 (92%; CI 0.82– 0.97) pregnant WWE remained seizure-free.17 One larger Class III article22 showed that 80% of a group of WWE (n _ 450) who were seizure-free at least 1 year prior to pregnancy remained seizure-free during pregnancy (exact number not provided). One Class III article showed that of 72 WWE who were seizure-free for 10 months, 74% (95% CI 0.62– 0.82) remained seizure-free during pregnancy.18 A second Class III article showed that of 54 WWE who were seizure-free for 9 months, 94% (95% CI 0.85– 0.98) remained seizure-free during pregnancy, and of 48 WWE who were seizure-free for 1 year, 92% (95% CI 0.80–0.98) remained seizure-free during pregnancy.19 These results are all fairly consistent across the class of evidence and sample size of the studies. *Conclusion.* Two Class II articles show the rate of remaining seizure-free during pregnancy if WWE are seizure-free for at least 9 months to 1 year prior to pregnancy is probably 84%–92%."

Harden CL, Hopp J, Ting TY, et al. Practice Parameter update: Management issues for women with epilepsy-Focus on pregnancy (an evidence-based review): Obstetrical complications and change in seizure frequency: Report of the Quality Standards Subcommittee and Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology and American Epilepsy Society. Neurology 2009;73:126-132.

AAN Management issues for women with epilepsy – focus on pregnancy guideline identified 9 studies graded Class III or higher. Regarding risk of reduced cognition, "Two Class II studies24,25 observed that cognition is not reduced in children of WWE unexposed to AEDs. One was a blinded observational study24 comparing the IQ of 64 children of WWE not

taking AEDs with 121 controls. No important differences in IQ were found. The other study25 showed no difference in the IQ of 57 children of untreated WWE and 57 control children matched for age, race, and socioeconomic status. ***Conclusion.*** Cognition is probably not reduced in children of WWE who are not exposed to AEDs in utero (two Class II studies)."

Regarding relationship between AED and risk of MCMs: "All studies evaluated AED dose in the first trimester and MCMs. In one Class I study,10 a relationship between AED dose and risk of MCMs was reported for LTG but not VPA. Using the Cochran Armitage method,17 we found a significant dose relationship with VPA (exact tests one-sided p _ 0.02, two-sided p _ 0.04) and with LTG (exact tests one-sided p _ 0.01, two-sided p _ 0.02), but not with CBZ (exact tests one-sided p _ 0.19, two-sided p _ 0.31). Two Class II studies11,12 and six Class III studies13-15,18-20 also found a relationship between VPA dose and MCMs. The VPA dose above which MCMs were significantly more likely to occur was not consistent, but was approximately 1,000 mg daily in five studies.12,13,18-20 **Are there specific MCMs associated with specific AEDs?** One Class I study10 showed increased risk of neural tube defects and facial clefts with VPA (RR 5.32, CI 1.38 –20.50 for neural tube defects and RR 4.18, CI 1.55–11.25 for facial clefts). One Class II study8 showed increased risk for cleft palate with PHT and posterior cleft palate with CBZ. Another

Class II study12 showed increased risk of neural tube defects and hypospadias with VPA. Two Class III studies showed increased risk of spina bifida with VPA,9,21 and one showed increased risk of hypospadias.

9 Two Class III studies9,15 showed increased risk of cardiac malformations associated with PB.

Conclusions

• PHT exposure in utero possibly contributes to the risk of cleft palate (one Class II study).

• CBZ exposure in utero possibly contributes to the risk of posterior cleft palate (one Class II study).

• VPA exposure in utero probably contributes to neural tube defects and facial clefts (one Class I study) and possibly contributes to hypospadias (one Class II study).

• PB exposure in utero possibly contributes to cardiac malformations (two Class III studies)."

Harden CL, Pennel PB, Koppel BS, et al. Practice Parameter update: Management issues for women with epilepsy-Focus on pregnancy (an evidence-based review): Vitamin K, folic acid, blood levels and breastfeeding: Report of the Quality Standards Subcommittee and Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology and American Epilepsy Society. Neurology 2009;73:142-149.

NICE guideline statements were a reaffirmation of previous statements in 2004. These statements and evidence were previously reviewed during prior endorsement review.

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

See details provided in 1a.7.5

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

See details provided in 1a.7.5

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

See details provided in 1a.7.5

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

NICE completed an update of the evidence in 2014, and complete details of their update can be found online at: http://www.nice.org.uk/guidance/cg137/evidence/evidence-update-544389949 NICE found there was no potential impact on guidance for Women and girls with epilepsy previously published.

- Vajda FJ, O'Brien TJ, Lander CM, et al. Tertogenesis in repeated pregnancies in antiepileptic drug-treated women. Epilepsia 2013; 54(1):181-186.
- Campbell E, Devenney E, Morrow J et al. (2013) Recurrence risk of congenital malformations in infants exposed to antiepileptic drugs in utero. Epilepsia 54: 165–71

- Meador KJ, Baker GA, Browning N et al. (2013) Fetal antiepileptic drug exposure and cognitive outcomes at age 6 years (NEAD study): a prospective observational study. Lancet Neurology 12: 244–52

_____

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
WWE_MeasSubm_Evidence_2016-635883584442931941.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
From original NQF submission -
Educate women about epilepsy and how its treatments may affect contraception and pregnancy. This will inform women of childbearing potential about the risks of epilepsy and AED therapy prior to pregnancy. It will provide an opportunity to educate this population about folic acid supplementation, monotherapy, medication alternatives and how to obtain obstetrical, prenatal and pregnancy care. This measure will help them understand the risk and mitigate the risks which may prevent fetal malformation, unplanned pregnancies and improve the patients´ quality of life.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
See Full AAN Women with Epilepsy of Childbearing Potential Measure Testing results provided in appendix.

Since the most recent endorsement, the measure has been used in CMS´s Physician Quality Reporting System (PQRS), and most recent PQRS data made available for the 2013 reporting period provided percentage of eligible neurologists participating, but did not provide more detailed information on performance by reporting neurologists on the Counseling for Women of Childbearing Potential with Epilepsy measure.
NeuroPI data records completion of module, and does not store or reconcile information on performance of specific measures in the epilepsy module. To date 198 physicians have completed the module out of 615 participants who have enrolled in the module.

Data collection through the AAN´s Axon Registry was initiated in Q3 of 2015. Data validation is ongoing, and further analysis of performance is not available at this time.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
In 2013, the AAN tested its Women with Epilepsy of Childbearing potential measure and evidence of a gap in care remains. Data from the testing project showed that on average less than 40% of women received counseling about epilepsy and how its treatment may affect contraception and pregnancy.(5) Additionally, the QUality Indicators for Epilepsy Treatment in adults (QUIET) study demonstrated that only 34% of female patients receive counselling on aspects of epilepsy care specific to women (neurologist alone=32.88%; shared (neurologists and primary care=44.83%; and primary care alone=11.11%).(6)

For babies whose mothers take seizure medication, the risk of birth defects is 4% to 8% compared with 2% to 3% for all babies.(7) Despite the availability of practice guidelines, knowledge about the use of seizure medications during pregnancy was low with less than half of neurologists able to identify which medications were linked to adverse events during pregnancy.(8)

(5) MN Community Measure, Women with Epilepsy Draft Testing Report. December 18, 2013.
(6) Pugh MJ, Berlowitz DR, Rao JK, et al. The quality of care for adults with epilepsy: an initial glimpse using the QUIET measure. BMC Health Services Research 2011;11:1. Available at: http://www.biomedcentral.com/1472-6963/11/1 Accessed on February 25, 2014.
(7) Epilepsy Foundation. Pregnancy issues website. Available at: www.epilepsyfoundation.org/living/women/pregnancy/weipregnancy.cfm. Accessed on February 25, 2014.

(8) Roberts, JI, Metcalfe A, Abdulla F, et al. Neurologists' and neurology residents' knowledge of issues related to pregnancy for women with epilepsy. Epilepsy Behav. 2011;22(2):358-363.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
See Full AAN Women with Epilepsy of Childbearing Potential Measure Testing results provided in appendix.

Since the most recent endorsement, the measure has been used in CMS´s Physician Quality Reporting System (PQRS), and most recent PQRS data made available for the 2013 reporting period provided percentage of eligible neurologists participating, but did not provide more detailed information on performance by reporting neurologists on the Counseling for Women of Childbearing Potential with Epilepsy measure.
NeuroPI data records completion of module, and does not store or reconcile information on performance of specific measures in the epilepsy module. To date 198 physicians have completed the module out of 615 participants who have enrolled in the module.

Data collection through the AAN´s Axon Registry was initiated in Q3 of 2015. Data validation is ongoing, and further analysis of performance is not available at this time.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
From original NQF submission -
A study in the Harlem neighborhood of New York City found epilepsy prevalence to be higher in Hispanics than in non-Hispanics and a higher prevalence of active epilepsy in whites than in blacks, although the prevalence of lifetime epilepsy was higher in blacks compared to whites (Kelvin et al., 2007). In this community, there were racial and ethnic disparities in care; blacks were more likely to receive care in the emergency department compared to whites and Hispanics. Similarly, Hope and colleagues (2009) found that blacks and Hispanics were more likely than whites to be diagnosed in an emergency department, and blacks were more likely to receive a suboptimal seizure medication. Differences in care for prevalent epilepsy were also observed in residents of Alabama and surrounding states, where blacks were 60 percent less likely than non-Hispanic whites to undergo epilepsy surgery after receiving electroencephalograph (EEG) monitoring as part of a surgical evaluation, an association that persisted after controlling for factors such as SES and medical insurance coverage (Burneo et al.,2005). The degree to which differences in epilepsy incidence and prevalence in different racial and ethnic groups reflect differences in socioeconomic status is unknown. Also unknown is the degree to which the treatment gap contributes to the higher epilepsy prevalence in some subgroups.

Kelvin EA, Hesdorffer DC, Bagiella E, et al. Prevalence of self-reported epilepsy in a multiracial and multiethnic community in New York City. Epilepsy Research. 2007;77(2-3):141-150.
Hope OA, Zeber JE, Kressin NR, et al. New-onset geriatric epilepsy care: Race, setting of diagnosis, and choice of antiepileptic drug. Epilepsia 2009; 50(5): 1085-1093.
Burneo JG, Black L, Knowlton RC, et al. Racial disparities in the use of surgical treatment for intractable temporal lobe epilepsy. Neurology. 2005; 64(1):50-54.

1c. High Priority (previously referred to as High Impact)
The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
From original NQF submission -
Recent estimates of the US population and prevalence of epilepsy indicate that approximately one-half million women with epilepsy

(WWE) are of childbearing age. It has also been estimated that 3-5 births per thousand will be to WWE. Epilepsy is defined by the presence of recurrent, unprovoked seizures, and the treatment is typically a daily, long-term antiepileptic drugs (AED) regiment. The majority of people with epilepsy have well-controlled seizures, are otherwise healthy, and therefore expect to participate fully in life experiences, including childbearing. Epilepsy is associated with sexual dysfunction, reduced fertility, increased pregnancy risks, and risks for malformations in the infant. Seizures can transiently disrupt pituitary hormone secretion. Treatment of seizures with antiepileptic drugs may alter hormone levels, render oral contraceptives less effective and may interfere with embryonic and fetal development. Certain antiepileptic mediations may have specific malformation risks. Since unplanned pregnancy is common, patients need to be informed about the risks of epilepsy and antiepileptic drug therapy prior to pregnancy. Folic acid supplementation, monotherapy for epilepsy, using lower doses of medication when possible and proper obstetrical, prenatal and pre-pregnancy care all should be discussed with the patient so they understand the risks involved and how to mitigate these risks.

The specific knowledge needed by women with epilepsy, which may vary by age, has generally received insufficient attention. Because sex hormones can affect seizure frequency, girls and women need information related to hormonal fluctuations and seizure frequency. Studies have found higher-than-expected onset of seizures during the year of menarche; in girls with preexisting seizures, 29 percent experienced more frequent seizures during perimenarche (Klein et al., 2003). Because of hormonal fluctuation, some women have a cyclic pattern of seizure frequency associated with their menses that often is unrecognized (Pennell and Thompson, 2009).

"Our suspicions have been confirmed: epilepsy affects women differently. Their hormonal and menstrual cycles, pregnancy, menopause—all of those life stages are affected by epilepsy," said Edna Kane-Williams, vice president of programs and services for the Epilepsy Foundation. Furthermore, Ms. Kane-Williams said, many medical professionals seem to be in the dark. "We´ve done a professional awareness survey that showed that the physicians these women were seeing weren´t aware of the differences," she reported. When women with epilepsy have problems, they are often hormone-based, according to Dr. Mark Yerby, founder of North Pacific Epilepsy Research in Portland, Oregon, and a nationally recognized authority on the subject.

Risks from seizures and from anti-epileptic drugs
Both seizures and medications are associated with some risks. The risk of seizures is associated with seizure type. Partial seizures probably do not carry as much risk but they may become generalized seizures, and generalized tonic-clonic seizures are associated with increased risk to both the mother and baby. These risks include trauma from falls or burns, increased risk of premature labor, miscarriages, and fetal heart rate suppression. Seizure control is necessary because the risks from seizures are felt by epileptologists to be greater than the risks from medications, which may be minimized by utilizing specific strategies.

Strategies to minimize risks
Most importantly, women should get accurate information prior to and during pregnancy. If anti-epileptic drugs are not needed, multiple medications are being taken, or medications are given at high dosages, changes should be considered with a neurologist prior to a planned pregnancy. The lowest possible anti-epileptic drug dose that will continue to maintain seizure control is recommended. Being on a single drug, monotherapy, will decrease the risk of birth defects and result in fewer drug interactions, fewer side effects, and improve compliance.

The 2012 IOM report “Epilepsy Across the Spectrum” explicitly stated the need for the development and implementation of a national quality measurement and improvement strategy for epilepsy care. “An independent organization with expertise in quality measurement and care should assist in the development of the national strategy, particularly the development of performance metrics.” Specifically, the IOM report calls for the national quality strategy to include defining performance metrics for epilepsy with specific attention to access to care for underserved populations, access to specialized care, co-management of care among specialized epilepsy providers, and coordination of care with other health care providers and community services organizations.

The AAN is a non-profit professional association with extensive experience and expertise in developing quality measures for neurological conditions and has developed eight quality measures for epilepsy care. The AAN has not yet completed testing of these measures. Three of the epilepsy measures were chosen for inclusion in the 2012 PQRS program and thus are under consideration for endorsement by the NQF at this time.

The incidence in females, at 41 cases per 100,000 person years, is less than that for males, at 49 cases per 100,000 person years.[2] Approximately 1 million women of childbearing age in the United States have seizure disorders. Of these women, approximately 20,000 give birth each year. Concerns during these pregnancies include the risk of fetal malformation, miscarriage, perinatal death, and increased seizure frequency.[1]
In women who are pregnant, the volume of distribution and the hepatic metabolism of AEDs are increased. This, along with decreased compliance with AEDs because of concerns about their effects on the fetus, leads to an increase in seizure frequency,

which is observed in as many as 17-33% of pregnancies.
The use of antiepileptic drugs (AEDs) is associated with a greater baseline risk of fetal malformations during pregnancy. When treating pregnant women who have epilepsy, the risks of increased seizure frequency versus the risks of AED use must be weighed carefully.
A population-based study conducted in Norway found that pregnant women with epilepsy had a lower risk of complications but an increased risk of induction, cesarean delivery, and postpartum hemorrhage.[2] However, whether this is a result of AEDs or severe epilepsy is unclear.

1c.4. Citations for data demonstrating high priority provided in 1a.3

United States Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, Bridged-Race Population Estimates, United States. July 1st resident population by state, country, age, sex, bridged-race, and Hispanic origin on CDC WONDER on-line Database. Available at: http://wonder/cdc.gov/ Accessed June 2012.

Hirtz D, Thurman DJ, Gwinn-Hardy K, et al. How common are the "common" neurological disorders? Neurology 2007; 68:326-337.

Yerby MS. Quality of life, epilepsy advances, and the evolving role of anticonvulsants in women with epilepsy. Neurology 2000; 55:S21-31.

Harden CL, Hopp J, Tin TY, et al. Practice parameter update: Management issues for women with epilepsy-Focus on pregnancy (an evidence-based review): Obstetrical complications and change in seizure frequency: Report of the Quality Standards Subcommittee and Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology and American Epilepsy Society. Neurology 2009; 73:126-132.

Practice parameter: management issues for women with epilepsy (summary statement). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 51(4):944-8, 1998

Pennell PB. The importance of monotherapy in pregnancy. Neurology. 60(11 Suppl 4):S31-8, 2003

Yerby MS. Management issues for women with epilepsy: neural tube defects and folic acid supplementation. Neurology. 61(6 Suppl 2):S23-6, 2003
Crawford, P., and S. Hudson. 2003. Understanding the information needs of women with epilepsy at different lifestages: Results of the "Ideal World" survey. Seizure 12(7):502-507.

Pennell, P. B., and P. Thompson. 2009. Gender-specific psychosocial impact of living with epilepsy. Epilepsy and Behavior 15(Suppl. 1):S20-S25

Shafer, P. O. Counseling women with epilepsy. Epilepsia 1998; 39(Suppl. 8):S38-S44.
2009. Epilepsy self-management in clinical practice: What we do and know. Paper read at AES Annual Meeting, Boston, MA: Hynes Conference Center.

Cramer JA, Gordon J, Schachter S, Devinsky O, and the Epilepsy Therapy Development Project Women's Issues Work Group. Women with Epilepsy: Hormonal Issues from Menarche through Menopause. Epilepsy Behav. 2007; 11: 160-178

Epilepsy Therapy Project
http://www.epilepsy.com/INFO/WOMEN_PREGNANCY

Institute of Medicine Report "Epilepsy Across the Spectrum: Promoting Health and Understanding"
http:www.iom.edu/Reports/2012/Epilepsy-Across-the-Spectrum.aspx

1. Katz O, Levy A, Wiznitzer A, Sheiner E. Pregnancy and perinatal outcome in epileptic women: a population-based study. J Matern Fetal Neonatal Med. Jan 2006;19(1):21-5.
2. Borthen I, Eide MG, Daltveit AK, Gilhus NE. Delivery outcome of women with epilepsy: a population-based cohort study. BJOG. Nov 2010;117(12):1537-43

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input*

was obtained.)
Not Applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology, Perinatal and Reproductive Health, Prevention

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Patient and Family Engagement, Prevention, Safety, Safety : Medication Safety

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
https://www.aan.com/practice/quality-measures/

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:** 1814_MeasSubm_MeasTesting_2016_01_27.docx

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** 2015-10-13_Epilepsy_Measure_6.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Measure numerator and denominator exceptions were updated for added specificity in response to testing data.

The numerator statement previously stated, "Female patients counseled about epilepsy and how its treatment may affect contraception and pregnancy and documented in the medical record at least once a year." Numerator statement provides explicit requirements for counseling to be met in order to achieve performance.

Prior version of measure exclusion stated, "Documentation of medical reason for not counseling the patient about epilepsy and how its treatment may affect contraception and pregnancy (e.g. patient is surgically sterile)." Specific medical reasons were detailed in this update.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Female patients or caregivers counseled* at least once a year about how epilepsy and its treatment may affect contraception OR pregnancy.

*Counseling should include a discussion about folic acid supplementation, contraception, potential anti-seizure medications effect(s) on pregnancy, safe pregnancies, and breastfeeding.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Within a 12 month reporting period

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of*

individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Measure is no longer applicable to claims reporting, and values (SNOMED) for gathering information through EHR and Registry sources is attached in S.2b.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All females of childbearing potential (12-44 years old) with a diagnosis of epilepsy.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Maternal Health

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Measure is no longer applicable to claims reporting, and values (SNOMED) for gathering information through EHR and Registry sources is attached in S.2b.

Epilepsy ICD-9-CM diagnosis codes
345.00, 345.01, 345.10, 345.11, 345.40. 345.41, 345.50, 345.51, 345.60, 345.61, 345.70, 345.71, 345.90, 345.91

OR Epilepsy ICD-10-CM diagnosis codes
G40.A09, G40.A19, G40.309, G40.411, G40.209, G40.219, G40.109, G40.119, G40.822, G40.824, G40.909

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Excluded: patients diagnosed with menopause or surgically sterile.

Exceptions:
Patient has a diagnosis of neurodevelopmental disorder, encephalopathy, hydrocephalus, brain injury, or cerebral palsy.

Patient has a diagnosis of severe cognitive impairment or severe intellectual disability.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Measure is no longer applicable to claims reporting, and values (SNOMED) for gathering information through EHR and Registry sources is attached in S.2b.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not Applicable

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
N/A

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
See Full AAN Women with Epilepsy of Childbearing Potential Measure Testing Report. MNCM completed validation of the data in a three-step process: 1) denominator certification, 2) data file quality checks, and 3) validation audit. Details of this validation are described in this report.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not Applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
For use in the American Academy of Neurology (AAN) Axon Registry only - If data is missing from denominator, the case is deleted. If data met for denominator then case information is included for the measure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Electronic Health Record, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
NeuroPI Program: for maintenance of certification Performance in Practice module.
CECity PQRSWizard
Axon Registry
Physician Quality Reporting System measurement set

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not Applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
1814_MeasSubm_MeasTesting_2016_01_27-635894858344728223.docx

Measure Number (*if previously endorsed*)**:** 1814

Measure Title: Counseling for Women of Childbearing Potential with Epilepsy

Date of Submission: 1/26/2016

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure*. *Testing must be provided for all the sources of data specified and intended for*

measure implementation. ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☒ abstracted from paper record	☒ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
The AAN identified and recruited Neurology practices in Minnesota that have experience treating patients with epilepsy. Three Neurology practices volunteered to participate and submit retrospective data from the 2012 calendar year (i.e. dates of service 01/01/2012 – 12/31/2012). Denominator certification is an essential step in the process to obtaining valid and accurate data. It requires each participant to attest that they will submit accurate data and follow the measure specifications exactly how they are written. It also ensures that each participant is querying the correct:

- Diagnosis codes *(i.e. 345.00, 345.01, 345.10, 345.11, 345.40, 345.41, 345.50, 345.51, 345.70, 345.71, 345.90, 345.91)*
- Encounter codes (*i.e. 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99241, 99242, 99243, 99244, 99245)*
- Date of birth ranges *(i.e. 01/01/1968-01/01/2000)*
- Date of service ranges *(i.e. 01/01/2012- 12/31/2012)*

1.3. What are the dates of the data used in testing? 01/01/2012-12/31/2012

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☒ group/practice	☒ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

See full testing report. Two of the practices were independent, physician owned practices located in the Twin Cities and the third practice was a large integrated delivery system in southeast Minnesota. One practice extracted all data via EHR; two practices had data extracted via EHR and also required manual chart review.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*) In order to achieve a reliable

sample of patients for this measure MNCM and AAN sought a minimum of 1,000 combined patient records from the three Neurology sites that agree to participate. See below.

Table 1: Patient Payer Information

Payer Type	Clinic A	Clinic B	Clinic C	Grand Total
Medicaid		76		76
Medicare	105	72	7	184
Self-Pay/Uninsured	9	11	2	22
Commercial	461	422	45	928
Medicaid	176		5	181

Table 2: Patient Place of Residence (based on zip code)

State	Group A Patients	Group B Patients	Group C Patients	Total
Minnesota	32	649	561	1242
Iowa	7	19	3	29
Wisconsin	4	49	14	67
North Dakota		9		9
South Dakota	3	11	1	15
Alaska		1		1
Arkansas	1			1
Colorado	1	1		2
Illinois	1	1		2
Maryland		1	1	2
Michigan	2	1		3
Nebraska	1	5	1	7
New York		1		1
North Carolina	1			1
Oklahoma	2			2
Ohio		1		1
Pennsylvania	1			1
Blank	3	2		5

Table 3: Patient Race and Ethnicity Information

Race	Group A Patients	Group B Patients	Group C Patients	Total
American Indian/Alaska Native (Code 1)		6	3	9
Asian (Code 2)	2	12	8	22
Black/African American (Code 3)		19	18	37
Hispanic/Latino (Code 4)		12		12
Native Hawaiian/Other Pacific Islander (Code 5)		3		3

Table 4: Patient Age Breakdown

Total by clinic	Group A Patients	Group B Patients	Group C Patients	Total
Ages 12-17	14	170	112	296
Ages 18-25	13	206	175	394
Ages 26-30	8	129	113	250
Age 31-35	13	111	73	197
Ages 36-40	8	93	60	161
Ages 41-44	3	42	48	93

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)
The AAN identified and recruited Neurology practices in Minnesota that have experience treating patients with epilepsy. As part of the recruitment process MNCM and the AAN hosted an informational webinar explaining the purpose of the measurement testing project for the Counseling for Women with Epilepsy measure. Three Neurology practices volunteered to participate and submit retrospective data from the 2012 calendar year (i.e. dates of service 01/01/2012 – 12/31/2012). MNCM produced a data collection guide, measure flow and detailed file specifications to educate and assist each medical group in the data collection and submission process. As a requirement of participating in the measure testing each group had to submit a denominator certifications form (see appendix A). The denominator certification process helps ensure that each medical group is using the appropriate measure parameters and collecting data in a standardized way.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)
MNCM did not identify any major flaws or issues during the review of each medical group's denominator forms and therefore each medical group passed denominator certification within the given timeframe.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)
There were a few corrections and clarifications that required MNCM to send a follow-up email to the respective group; however, each issue was resolved in a timely manner. The list below documents the issues that were identified and required additional follow-up based on the information received on the denominator certification forms:

- Incorrect diagnosis codes included in data query
- Group did not indicate if they would be submitting a sample or full population for the measure
- Incorrect encounter codes included in data query

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*
MNCM completed validation of the data in a three-step process: 1) denominator certification, 2) data file quality checks, and 3) validation audit. Details of this validation are described in this report.

Denominator Certification
Denominator certification is an essential step in the process to obtaining valid and accurate data. It requires each participant to attest that they will submit accurate data and follow the measure specifications exactly how they are written. It also ensures that each participant is querying the correct:

- Diagnosis codes *(i.e. 345.00, 345.01, 345.10, 345.11, 345.40, 345.41, 345.50, 345.51, 345.70, 345.71, 345.90, 345.91)*
- Encounter codes (*i.e. 99201, 99202, 99203, 99204, 99205, 99212, 99213, 99214, 99215, 99241, 99242, 99243, 99244, 99245)*
- Date of birth ranges *(i.e. 01/01/1968-01/01/2000)*

- Date of service ranges *(i.e. 01/01/2012- 12/31/2012)*

MNCM did not identify any major flaws or issues during the review of each medical group's denominator forms and therefore each medical group passed denominator certification within the given timeframe. There were a few corrections and clarifications that required MNCM to send a follow-up email to the respective group; however, each issue was resolved in a timely manner. The list below documents the issues that were identified and required additional follow-up based on the information received on the denominator certification forms:

- Incorrect diagnosis codes included in data query
- Group did not indicate if they would be submitting a sample or full population for the measure
- Incorrect encounter codes included in data query

Data File Quality Checks

After each medical group submitted their data file to MNCM, quality checks of the files were completed. Each column in the data file represented a field of data for each patient row; the following checks were completed:

- Number of patients/rows submitted were reasonable/expected
- Necessary data fields (columns) were included and completed appropriately
- Patient date of birth spanned the expected range
- Zip codes were 5-digit and primarily within MN and other bordering states as expected
- Race field(s) were included and populated appropriately
- Provider NPI field was included and number of providers expected
- Insurance information was included and was reasonable
- Office visit dates and counseling dates spanned the expected range
- Diagnoses were included and spanned the entire list of expected codes
- Medical reasons for NOT counseling were applied correctly; were not misused

Issues identified through the data file quality checks were generally minor, requiring no corrections. Other mentionable items include:

1. All three groups did not have patients with diagnosis codes 345.70 (Epilepsia partialis continua, without mention of intractable epilepsy) or 345.71 (Epilepsia partialis continua, with intractable epilepsy). These are rare diagnoses and did not come up in the population.

2. Medical groups B and C listed many neurological or congenital conditions as reasons for the patient to NOT receive counseling. These were verified during audit.

3. Medical group C did not include their entire population in first submission, excluding patients whose date of birth was between January thru June 1968. They queried their system again, this time using the specific dates of birth (rather than "age" values) and included the additional patients in their denominator.

Validation Audit

After the data file checks were completed, MNCM completed audits of the patient records to verify the submitted clinical data. We also verified the diagnosis of epilepsy and other demographic data (e.g., race). MNCM uses a validation process developed by the NCQA – National Committee for Quality Assurance, known as the "8 and 30" process. In this process, the first eight records are verified for accuracy and if no errors are identified, the data is considered to be 100% compliant. If errors in the first eight records are identified, we continue reviewing the total 30 records to identify any error patterns and or issues that may need correction. The audits revealed some data errors, requiring one medical group to make corrections and resubmit data. Individual medical group results were as follows:

Medical group	Audit details	Follow-up action
A	8 records reviewed, 8 records compliant (100%) 3 additional records were reviewed for "other" reason patient was not counseled ○ 2 records were compliant ○ 1 record indicated patient had functional seizures and not epilepsy, but should have been counted as no counseling provided	No further action necessary
B	30 records reviewed, 26 compliant (87%) ○ Errors: three records had code "2" for no counseling due to intellectual disability, however, we could not verify the diagnosis in the record; one record reported as "1" counseling given could not be verified 3 additional records were reviewed whose patients were listed as "cognitively impaired" as a reason for not receiving counseling; verified that these patients had mild retardation; medical group staff corroborated that all 115 patients with this designation also had mild retardation	Group verified patients they submitted who were "surgically sterile" or who had "intellectual disability"; resubmitted data with corrections
C	8 records reviewed, 8 records compliant (100%) We identified one record in the eight reviewed in which the patient could have been flagged for a medical reason to NOT receive counseling (99 "other), but because the reasons were not either type (surgically sterile, intellectual disability), it was appropriate that these patients could have been counseled; these were not counted as errors	No further action necessary

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Validation/ Audit Conclusion

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)
The validation process was successful in identifying errors (with subsequent corrections) and verifying the accuracy of the data submitted by medical groups A, B, and C. Finding no significant flaws or errors with the data MNCM is confident the rate calculation and any additional data analysis can be completed using validated and reliable data. Additionally, during a review of the National Quality Forum's feedback to the American Academy of Neurology for this measure, it was noted that there was a concern that this may simply be a "check-the-box" measure. During the validation audit, it was noted on several occasions that the practices provided excellent, personalized progress notes about the counseling that was being provided, that were above and beyond a "check the box".

_____

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)
The main limitation that MNCM identified during the testing of the Counseling for Women with Epilepsy measure is related to the denominator of included and excluded patients. The measure specifications offered two different options for excluding patients from the measure:
1. Patient was surgically sterile (tubal ligation, hysterectomy)
2. Patient has an intellectual disability as defined by ICD-9 codes
a. 318.0 moderate intellectual disabilities; IQ 35 to 48
b. 318.1 severe intellectual disabilities; IQ 20 to 34
c. 318.2 profound intellectual disabilities; IQ under 20

Groups submitted these patients and indicated which reason applied. Additionally, if they felt that there was another medical reason for not providing counseling, they indicated this by a code and accompanying description. These reasons were not used to exclude patients from the measure; rather the purpose was to provide additional information about the population of patients included in the measure. Reasons Provided by Medical Groups for Not Providing Counseling:

Reason by Frequency	Count	Valid	Thoughts
cognitive impairment/ deficit	138	Maybe	subjective and may still be at risk
cerebral palsy	45	Yes	need to quantify by code
neurodevelopmental disorder	18	Yes	need to quantify by code
encephalopathy	15	Yes	need to quantify by code
developmental delay	14	No	may still be at risk
hydrocephalus	8	Yes	need to quantify by code
brain injury	8	Yes	need to quantify by code
pregnancy	7	No	still needs counseling
pre-menarche	7	No	may still be at risk
autism	4	No	may still be at risk
downs syndrome	3	Maybe	may still be at risk
aspergers	2	No	spectrum of functioning; at risk
birth control- IUD	2	No	may still be at risk
learning disability	2	No	spectrum of functioning; at risk
menopause	1	Yes	need to quantify by code
no menses	1	Maybe	
multiple sclerosis	1	No	Spectrum of functioning; at risk

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Measure Results

Measure Results	Clinic A	Clinic B	Clinic C	Total
Number of Providers (NPI)	9	36	22	67
Number of Patients Submitted	751	581	59	1391
Number of Patients Excluded; intellectual disability codes	127	32	2	161
Number of Patients Excluded; surgically sterile	18	12	5	35
Number of Patients with valid exclusions	145	44	7	196
Denominator: Number of Patient Eligible for Counseling	606	537	52	1195
Number of Patients with Counseling for Contraception	420	26	11	457
Number of Patients with Counseling for Pregnancy	419	77	7	503
Number of Patients with Contraception and Pregnancy	419	21	6	446
Rate for Contraceptive Counseling	**69.3%**	**4.8%**	**21.2%**	**38.2%**
Rates for Pregnancy Counseling	**69.1%**	**14.3%**	**13.5%**	**42.1%**
Rates for Contraceptive and Pregnancy Counseling	**69.1%**	**3.9%**	**11.5%**	**37.3%**

Other Medical Reasons for Not Counseling Patients	Clinic A	Clinic B	Clinic C	Total
Number of Patients Submitted	751	581	59	1391
Number of Patients with "Other Medical Reason"	122	156	6	446
Percentage of Patients with "Other Medical Reason"	16.2%	26.9%	10.2%	20.4%
Please note: The following rate re-calculation is for analytical purposes only; removing all patients that had "Other Medical Reason Documented". MNCM does not recommend reporting this rate. Please see Limitations Section				
Rates for Contraceptive and Pregnancy Counseling if these patients are also removed from the denominator	86.6%	5.5%	13.0%	49.0%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)
During the analysis of this data and also as a byproduct of the validation audit in reviewing medical records, MNCM staff has some concerns regarding the denominator and intent of the measure. There may need to be a consideration for adding a component of indicating that the patient is sexually active or has the potential to be sexually active, and not physically handicapped. AAN could refer to the NCQA specifications for the Chlamydia Screening in Women measure (NQF# 0033/ CMS 153v1) for reference tables indicating how to identify potentially sexually active women via pharmacy codes, CPTs, ICD-9, UB Revenue and LOINC codes. Rather than trying to capture/ code every possible exclusion; this may be an option. MNCM would not recommend having a general type exclusions code, like one that is stated as "any documented medical reason, because providers will use this to their advantage and exclude patients that are at risk for pregnancy and truly belong in the denominator. Having this type of exclusion weakens the measure, and can impact the validity and reliability of the results.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care and not related to disparities*)

2b4.4. What were the statistical results of the analyses used to select risk factors?

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

_____

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i*.e., what do the results mean and what are the norms for the test conducted*)

_____

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

See Full AAN Women with Epilepsy of Childbearing Potential Measure Testing Report. The validation process is conducted to verify that the submitted data matches the source data in the medical record. After the clinical data file is successfully transferred to MNCM and passes the initial quality checks, MNCM will contact the medical group about the validation process. MNCM utilizes the National Committee for Quality Assurance (NCQA) "8 and 30" process for validation audits. If the first clinic site is in high compliance and the data collection process for all clinic sites within the medical group is identical, further review may be abbreviated at the discretion of the MNCM auditor. If clinic sites are not in high compliance after review of the first eight records, the MNCM auditor will continue to review the remaining 22 records. If after review of all 30 records the clinic site is not in high compliance on all factors (less than 90%), the MNCM auditor will review the results with the clinic representative and communicate the results with MNCM. MNCM will then contact the medical group to develop a mutually agreed upon re-submission plan. (Re-submission plans will only be allowed for errors in the numerator portion.) Clinic sites that are not in high compliance or have not been in high compliance in a previous MNCM audit may be held to a more rigorous denominator certification and validation audit.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i*.e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

MNCM did not identify any major flaws or issues during the review of each medical group's denominator forms and therefore each medical group passed denominator certification within the given timeframe. There were a few corrections and clarifications that required MNCM to send a follow-up email to the respective group; however, each issue was resolved in a timely manner. The list below documents the issues that were identified and required additional follow-up based on the information received on the denominator certification forms:

- Incorrect diagnosis codes included in data query
- Group did not indicate if they would be submitting a sample or full population for the measure
- Incorrect encounter codes included in data query

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in a combination of electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

No feasibility assessment **Attachment:**

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Based on MNCM testing additional exclusion specification were included in this update. The AAN´s Axon Registry was initiated in 2015 Q3 and lessons are ongoing. Data is being successfully pulled from EHRs without need for manual extraction. This information will be used to strengthen the measure in the next scheduled update in Q4 2016.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

Not applicable

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Payment Program PQRS https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/index.html?redirect=/pqri/ Professional Certification or Recognition Program NeuroPI http://tools.aan.com/practice/pip/ Quality Improvement (Internal to the specific organization) Axon Registry https://www.aan.com/practice/axon-registry/

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

NeuroPI Maintenance of Certification program by the AAN Available at: http://tools.aan.com/practice/pip/

The American Academy of Neurology Axon Registry. Currently an internal benchmarking quality registry with plans to expand to external benchmarking. This tool enables neurology practices to identify and improve gaps in the quality of neurologic care. Axon Registry was launched in Q3 of 2015. The Women with Epilepsy measure will be incorporated into the registry in 2016. Currently there are 3 pilot cohorts in progress providing data to the registry. Axon Registry is currently being piloted and will be available to all AAN US members. Currently, there are 42 practices (487 neurology providers) in the 3 pilot cohorts.

CMS Physician Quality Reporting System (PQRS) 57.5% of eligible neurologists participated in PQRS reporting in 2013 (last known data). Data on this measure not known, as not reported in the PQRS and ERX Experience Report. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/2013_PQRS_eRx_Experience_Report_zip.zip

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Not Applicable

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Not Applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

See Full AAN Women with Epilepsy of Childbearing Potential Measure Testing results provided in appendix.

Since the most recent endorsement, the measure has been used in CMS´s Physician Quality Reporting System (PQRS), and most recent PQRS data made available for the 2013 reporting period provided percentage of eligible neurologists participating, but did not provide more detailed information on performance by reporting neurologists on the Counseling for Women of Childbearing Potential with Epilepsy measure.

NeuroPI data records completion of module, and does not store or reconcile information on performance of specific measures in the epilepsy module. To date 198 physicians have completed the module out of 615 participants who have enrolled in the module.

Data collection through the AAN´s Axon Registry was initiated in Q3 of 2015. Data validation is ongoing, and further analysis of performance is not available at this time.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Unable to assess at this time.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
No unintended negative consequences have been identified since implementation of updated specifications.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;

OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Not Applicable

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not Applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Women_With_Epilepsy_Supporting_Materials.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Academy of Neurology
Co.2 **Point of Contact:** Amy, Bennett, abennett@aan.com, 612-928-6072-
Co.3 **Measure Developer if different from Measure Steward:** American Academy of Neurology
Co.4 **Point of Contact:** Amy, Bennett, abennett@aan.com, 612-928-6072-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Voting Work Group Members:
Nathan Fountain, MD (Chair for American Academy of Neurology)
Paul C. Van Ness, MD (Chair for American Academy of Neurology)

American Academy of Neurology
Jerome Engel, Jr., MD, PhD, FAAN
David S. Gloss, MD
Christiane Heck, MD, MMM
Diego A. Morita, MD
Marianna V. Spanaki, MD, PhD, MBA
Thaddeus Walczak, MD
American Academy of Family Physicians
Mark C. Potter, MD
American Academy of Pediatrics

Edwin Trevathan, MD, MPH
American Association of Neurological Surgeons/Congress of Neurosurgeons
Joseph Neimat, MD
American Association of Neuroscience Nurses
Mona Stecker, DNP, NP-BC, CNRN, SCRN
American Board of Internal Medicine
Sharon M. Hibay, RN, DNP
American Clinical Neurophysiology Society
Susan T. Herman, MD
American College of Emergency Physicians
J. Stephen Huff, MD
American Epilepsy Society
Gabriel U. Martz, MD
American Society of Neuroradiology/American College of Radiology
Marvin Nelson, MD
Child Neurology Society
Inna Hughes, MD, PhD
Citizens United for Research in Epilepsy
Tracy Dixon-Salazar, PhD
Epilepsy Foundation
Janice M. Buelow, RN, PhD
National Academy of Neuropsychology
Daniel Drane, PhD, ABPP(CN)
National Association of Epilepsy Centers
Ramon Bautista, MD, MBA
OptumInsight
Kay Schwebke, MD, MPH, MA
Veterans Affairs Epilepsy Centers of Excellence
Karen Parko, MD, FAAN
Independent Representatives
Laurie A. Olmon
Mary Jo Pugh, PhD, RN

Non-Voting Work Group Members:
John R. Absher, MD, FAAN (Facilitator)
Anup D. Patel, MD (Facilitator)
Kevin N. Sheth, MD, FAHA, FCCM, FNCS (Facilitator)

Amy Bennett, JD (AAN Staff)
Gina Gjorvad (AAN Staff)
Becky Schierman, MPH (AAN Staff)
Rebecca J. Swain-Eng, MS, CAE (AAN Staff)

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:** 2009
Ad.3 **Month and Year of most recent revision:** 08, 2014
Ad.4 **What is your frequency for review/update of this measure?** Every 3 years
Ad.5 **When is the next scheduled review/update for this measure?** 11, 2016

Ad.7 **Disclaimers:** Quality Measures (Measures) and related data specifications developed by the American Academy of Neurology (AAN) are intended to facilitate quality improvement activities by providers.

These measures are intended to assist providers in enhancing quality of care. Measures are designed for use by any provider who manages the care of a patient for a specific condition or for prevention. These Measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications. The AAN encourages testing and evaluation of its Measures.

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 1952
Measure Title: Time to Intravenous Thrombolytic Therapy
Measure Steward: American Heart Association/American Stroke Association
Brief Description of Measure: Acute ischemic stroke patients aged 18 years and older receiving intravenous tissue plasminogen activator (tPA) therapy during the hospital stay and having a time from hospital arrival to initiation of thrombolytic therapy administration (door-to-needle time) of 60 minutes or less.
Developer Rationale: Multiple studies have shown that the rapid administration of intravenous recombinant tissue-type plasminogen activator (tPA) to appropriate patients is a proven, effective treatment in restoring blood flow and reducing long-term morbidity for ischemic stroke patients.

With every 10-minute delay in the start of thrombolytic infusion within the 1-to-3 hour-treatment time period, there was 1 fewer patient of 100 patients having improved disability outcome. The administration of tPA is time-dependent and shortening door-to-needle time helps improve functional outcomes at 4 to 6 months when given in the recommended timeframe (8, 9). The Target Stroke Initiative to improve the timeliness of tPA administration found that among 71,169 patients with acute ischemic stroke that were treated with tPA at 1,030 GWTG stroke participating hospitals, participation in the program was associated with decreased door-to-needle time, lower in-hospital mortality and intracranial hemorrhage, and an increase in the percentage of patients discharged home (10). Results from clinical trials and registries have encouraged multiple organizations to set targets for timely initiation of thrombolytic therapy after hospital arrival. A National Institute of Neurological Disorders and Stroke national symposium on the rapid identification and treatment of acute stroke recommended a door-to-needle target time of 60 minutes (1). American Heart Association/American Stroke Association guidelines recommend the target for completion of initial evaluation and start of tPA treatment should be within 60 minutes of the patient´s arrival in the emergency department (2-3). The Brain Attack Coalition´s target for primary stroke centers is to achieve a door-to-needle time within 60 minutes in 80% or more of patients (4). Despite evidence and recommendations advocating a door-to-needle time of 60 minutes or less, studies have shown that less than 30% of US patients are treated within this time window (5-7). This measure is intended to promote a reduction in door to needle times and improvement in the proportion of eligible patients receiving treatment within 60 minutes of hospital arrival.

Citations:

1. Marler JR, Winters Jones P, Emr M. The National Institute of Neurological Disorders and Stroke: Proceedings of National Symposium on Rapid Identification and Treatment of Acute Stroke. Bethesda, MD: National Institute of Neurological Disorders and Stroke, 1997.

2. Adams HP Jr., del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, et al. Guidelines for the early management of adults with Ischemic stroke. Circulation. 2007;115:e478–534. Available at: http://circ.ahajournals.org/content/115/20/e478.full

3. Summers D, Leonard A, Wentworth D, Saver JL, Simpson J, Spilker JA, et al. Comprehensive overview of nursing and interdisciplinary care of the acute ischemic stroke patient: a scientific statement from the American Heart Association.

Stroke. 2009;40:2911–2944. Available at: http://stroke.ahajournals.org/content/40/8/2911.full.pdf+html

4. Alberts MJ, Hademenos G, Latchaw RE, Jagoda A, Marler JR, Mayberg MR, et al. Recommendations for the establishment of primary stroke centers. Brain Attack Coalition. JAMA. 2000;283:3102–3109. Available at: http://jama.ama-assn.org/content/283/23/3102.full

5. Fonarow GC, Zhao X, Smith EE, Saver JL, Reeves MJ, Bhatt DL, et al. Door-to-needle times for tissue plasminogen activator administration and clinical outcomes in acute ischemic stroke before and after a quality improvement initiative. JAMA. 2014;311(16):1632-1640. DOI:10.1001/jama.2014.3203.

6. Jauch EC, Saver JL, Adams HP, Bruno A, Connors JJ, Demaerschalk, BM, et al; American Heart Association Stroke Council; Council on Cardiovascular Nursing; Council on Peripheral Vascular Disease; Council on Clinical Cardiology. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44(3):870-947.

7. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics- 2016 update a report from the American Heart Association. Circulation. 2015;132:000-000. DOI: 10.1161/CIR.0000000000000350.

8. Lansberg MG, Schrooten M, Bluhmki E, Thijs VN, Saver JL. Treatment time-specific number needed to treat estimates for tissue plasminogen activator therapy in acute stroke based on shifts over the entire range of the modified Rankin Scale. Stroke. 2009;40:2079 –2084. Available at: http://stroke.ahajournals.org/content/40/6/2079.full

9. Powers WJ, Derdeyn CP, Biller J, Coffey CS, Hoh BL, Jauch EC, et al. 2015 American Heart Association/American Stroke Association focused update of the 2013 guidelines for the early management of patients with acute ischemic stroke regarding endovascular treatment. Stroke. 2015;46: 3024-3039. DOI:10.1161/STR.0000000000000074

10. Fonarow G, Smith EE, Saver J, Reeves M, Bhatt D, Grau-Sepulveda M, Olson D, Hernandez A, Peterson E, Schwamm L. Timeliness of Tissue-Type Plasminogen Activator Therapy in Acute Ischemic Stroke: Patient Characteristics, Hospital Factors, and Outcomes Associated With Door-to-Needle Times Within 60 Minutes. Circulation. 2011, 123:750-758: originally published online February 10, 2011 doi: 10.1161/CIRCULATIONAHA.110.974675. Available at: Available at: http://circ.ahajournals.org/content/123/7/750.full.pdf

Numerator Statement: Acute ischemic stroke patients aged 18 years and older receiving intravenous tissue plasminogen activator (tPA) therapy during the hospital stay and having a time from hospital arrival to initiation of thrombolytic therapy administration (door-to-needle time) of 60 minutes or less.
Denominator Statement: All acute ischemic stroke patients who received intravenous thrombolytic therapy during the hospital stay.
Denominator Exclusions: Denominator Exclusions:
•Patients less than 18 years of age
•Patient stroke occurred while in hospital
•Patients received in transfer from the inpatient, or outpatient of another facility
•Patients that receive tPA greater than 4.5 hours after Last Known Well
•Clinical trial

Denominator Exceptions:
Patients with documented Eligibility or Medical reason for delay in treatment [eg, social, religious, initial refusal, hypertension requiring aggressive control with intravenous medications, inability to confirm patients eligibility, or further diagnostic evaluation to confirm stroke for patients with hypoglycemia (blood glucose < 50); seizures, or major metabolic disorders, or management of concomitant emergent/acute conditions such as cardiopulmonary arrest, respiratory failure requiring intubation), or investigational or experimental protocol for thrombolysis.]

Measure Type: Process

Data Source: Electronic Clinical Data : Registry
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Nov 01, 2012 **Most Recent Endorsement Date:** Nov 01, 2012

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

- Evidence for this process measure should demonstrate that when acute ischemic stroke patients aged 18 years and older receive intravenous tissue plasminogen activator (tPA) therapy during the hospital stay and have a time from hospital arrival to initiation of thrombolytic therapy administration (door-to-needle time) of 60 minutes or less will have lower in-hospital mortality and intracranial hemorrhage, better clinical and functional outcomes.

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary or **Summary of prior review in [year]**

Rapid administration of intravenous tPA treatment >timely restoration of blood flow in ischemic stroke patient > decrease in morbidity and mortality. Guideline was supported by multiple studies including several meta-analysis, RCT trials and observational studies.

- The AHA/ASA Guidelines (2007) for Early Management of Adults with Ischemic Stroke include a Class I/Level of Evidence B recommendation that "An organized protocol for the emergency evaluation of patients with suspected stroke is recommended. The goal is to complete an evaluation and to decide treatment within 60 minutes of the patient's arrival in an ED.
- The NINDS rt-PA Stroke Study demonstrated among 624 patients with ischemic stroke treated with placebo or rt-PA (0.9 mg/kg IV, maximum 90 mg) within 3 hours of symptom onset that favorable outcomes were achieved.
- An analysis of 25,504 ischemic stroke patients treated with tPA among 1,082 Get With The Guidelines–Stroke hospitals demonstrated lower in-hospital mortality and less frequent symptomatic intracranial hemorrhage for patients with door-to-needle times <=60 minutes compared with patients with door-to needle times >60 minutes.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**
☒ **The developer provided updated evidence for this measure:**

Updates:

- Developers presented the AHA/ASA 2013 Guideline for this measure in patients eligible for intravenous rtPA, benefit of therapy is time dependent, and treatment should be initiated as quickly as possible. Specifically the door-to-needle time (time of bolus administration) should be within 60 minutes from hospital arrival. (Class 1, Level of Evidence A).
- Guideline was based on 16 randomized control trials, 1 open trial, 32 observational studies, and 4 meta-analyses.
- These two additional studies (a meta-analysis and an observational study), both find that early administration of tPA is associated with better outcomes.

Exception to evidence

Not applicable

Guidance from the Evidence Algorithm

Process measure assesses performance, based on systemic review(Box 3) >→ QQC presented (Box 4) > AHA/ASA guidelines & meta-analysis: Quantity: high; Quality: high; Consistency high(Box 5) → Box 5a

Questions for the Committee:

- *The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF evaluation. Does the Committee agree that there is no need for repeat discussion and vote on Evidence?*

Preliminary rating for Evidence ☒ High ☐ Moderate ☐ Low ☐ Insufficient

Preliminary rating for evidence: ☒ Pass ☐ No Pass

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- Recent studies have indicated only 30% of patients receive tPA treatment within the guideline-recommended 60 minute door-to-needle time. (Fonarow et al, 2014; Mozaffarian et al., 2015). This indicates a substantial gap in the compliance with this measure
- Recent data from the developer shows that the mean performance score using data from the Get with the Guidelines registry increased from 53% to 70% (2012 to 2015).
- The developer presented data as specified stratified by age, sex, and race/ethnicity, with results indicating lower performance for women compared to men, and higher performance for some minorities but not others.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Preliminary rating for opportunity for improvement: ☒ High ☐ Moderate ☐ Low ☐ Insufficient

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **This is a process measure. This measure directly relates to improving the outcome of better function after ischemic stroke in that reducing the time to treatment from onset of stroke increases the likelihood of a good functional outcome with IV TPA infusion.

EVIDENCE:

Jauch EC, Saver JL, Adams HP Jr, et al; American Heart Association Stroke Council, Council on Cardiovascular Nursing, Council on

Peripheral Vascular Disease, and Council on Clinical Cardiology. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013 Mar;44(3):870-947

American College of Emergency Physicians, American Academy of Neurology. Clinical Policy: Use of intravenous tPA for the management of acute ischemic stroke in the emergency department. Ann Emerg Med. 2013 Feb;61(2):225-43

Harris D, Hall C, Lobay K, etc. Canadian Association of Emergency Physicians position statement on acute ischemic stroke. CJEM. 2015 Mar;17(2):217-26

Casaubon LK, Suddes M, on behalf of the Acute Stroke Care Writing Group. Chapter 3: Hyperacute Stroke Care. In Lindsay MP, Gubitz G, Bayley M, and Phillips S, eds, on behalf of the Canadian Stroke Best Practices and Standards Advisory Committee. Canadian Best Practice Recommendations for Stroke Care: 2013; Ottawa, Ontario Canada: Canadian Stroke Network and Heart and Stroke Foundation of Canada. Updated May 23, 2013. Canadian Stroke Best Practice Recommendations 2013 May 23

1b. Performance Gap

Comments: **Key practice gap is low rate of IV TPA use in acute ischemic stroke for various reasons. This measure does not address this performance gap directly, but may increase the likelihood of patients being treated within the 3-4.5 hour time window from onset by improving the component of door to needle time.

EVIDENCE: Cocho D, Belvis R, Marti-Fàbregas J, Molina-Porcel L, Diaz-Manera J, Aleu A,et al. Reasons for exclusion from thrombolytic therapy following acute ischemicstroke. Neurology 2005;64:719–20.[3] Johnston SC, Fung LH, Gillum LA, Smith WS, Brass LM, Lichtman JH, et al. Uti-lization of intravenous tissue-type plasminogen activator for ischemic strokeat academic medical centers: the influence of ethnicity. Stroke J Cereb Circul2001;32:1061–8.[4] Katzan IL, Hammer MD, Hixson ED, Furlan AJ, Abou-Chebl A, Nadzam DM,et al. Utilization of intravenous tissue plasminogen

1c. High Priority (previously referred to as High Impact)

Comments: ** None

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s):

Specifications:

- This is from a clinical registry (GWTG-Stroke).
- The level of analysis is at the hospital/facility level
- The denominator includes acute ischemic stroke patients who received intravenous thrombolytic therapy during the hospital stay.
- The numerator includes acute ischemic stroke patients aged 18 years and older receiving intravenous tissue plasminogen activator (tPA) therapy during the hospital stay and having a time from hospital arrival to initiation of thrombolytic therapy administration (door-to-needle time) of 60 minutes or less.
- The denominator exclusions include patients : less than 18 years of age; stroke occurred while in hospital; received in transfer from the inpatient, or outpatient of another facility; that receive tPA greater than 4.5 hours after Last Known Well; and those in a clinical trial.
- The denominator exceptions include patients: with documented eligibility or medical reason for delay in treatment.
- For data collection the Get with the Guidelines Stroke Data Collection Form is used. This is a paper version of the electronic data collection tool which is called the Patient Management Tool (PMT).

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*

2a2. Reliability Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

For maintenance measures, summarize the reliability testing from the prior review:

- A stroke module data quality audit was conducted in 2010 that consisted of 438 medical records randomly selected from 147 GWTG-Stroke hospitals, with records from each hospital containing at least one tPA patient. The data submitted to the GWTG-Stroke program were compared against the medical record by a trained coder at the independent statistical coordinating center. No significant differences among participating hospitals were found in overall Inter-rater reliability by bed size, ischemic stroke volume, primary stroke center certification, or Coverdell Registry participation.

Describe any updates to testing

- The developer conducted score level testing

- **SUMMARY OF TESTING**

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**
Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

Prior testing:

- Accuracy for each individual data element and a composite accuracy measure were calculated. Agreement was assessed using kappa (K) statistics for categorical variables and intraclass correlation (ICC) for continuous variables. NQF considers this to be data element validity testing and therefore additional reliability testing isn't required.

Updated method of testing:

- Empirical validity testing at the performance measure score level was conducted via a signal-to-noise analysis using the beta-binomial model. This is an appropriate method for testing reliability.
- A signal-to-noise analysis quantifies the amount of variation in performance that is due to differences between hospitals (as opposed to differences due to random measurement error). Results will vary based on the amount of variation between the hospitals and the number of patients treated by each hospital. This method results in a reliability statistic that ranges from 0 to 1 for each provider. A value of 0 indicates that all variation is due to measurement error and a value of 1 indicates that all variation is due to real differences in provider performance. A value of 0.7 is often regarded as a minimum acceptable reliability value.
- Data used for testing included information from 672 of the 841 hospitals (79.9%) that reported data on this measure to the GWTG-Stroke registry and had at least 10 "reporting events" between October 2014 and September 2015 (note that a patient theoretically could have more than one reporting event during the measurement period). These data included information on 16,100 patients.

- Developers computed two "overall" reliability statistics: one that would be achieved if all hospitals had 10 reporting events and one that would be achieved if all hospitals had 23.9 reporting events (note: 23.9 is the average number of reporting events across the 672 hospitals).

Results of reliability testing

Prior results:

- See prior validity testing, below.

This measure had a inter-rater reliability of 0.72 (kappa statistic, 95% confidence interval 0.37-1.00.) including reliability of Door to IV tPA <= 60 minutes (yes/no.) . According to the Landis & Koch classification, a kappa of 0.72 is interpreted as substantial agreement

Updated results:

- Reliability if all hospitals had 10 reporting events: 0.63
- Reliability if all hospitals had 23.9 reporting events: 0.81
- Of the 841 hospitals reporting to the registry , 20.1% had fewer than 10 reporting events, and thus were not included in the reliability testing (note that the measure is not specified to excluded low-volume hospitals). The developers do not indicate the number of hospitals with at least 23.9 reporting events.

Guidance from the Reliability Algorithm

Specifications are clear and precise(Box 1)>Empirical testing using statistical tests ((Box 2)>score-level testing (Box 4)-appropriate method used (signal to noise) (Box 5)>moderate confidence of reliability (Box 6b)

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

For maintenance measures, summarize the validity testing from the prior review:

Prior testing:

Data element testing:

- Agreement was assessed using kappa (K) statistics for categorical variables and intraclass correlation (ICC) for continuous variables. The data submitted to the GWTG-Stroke program were compared against the medical record by a trained coder at the independent statistical coordinating center. No significant differences among participating hospitals were found in overall Inter-rater reliability by bed size, ischemic stroke volume, primary stroke center certification, or Coverdell Registry participation.

Face validity:

- A review of the relevant evidence and guidelines and expert panel consensus process resulted in the conclusion that this is a valid measure of quality of stroke care for patients receiving an IV thrombolytic therapy within 60 minutes.

Describe any updates to testing:

Face Validity:

- Developer conducted a face validity assessment by 20 experts from the AHA Council on Stroke 2015-2016 Leadership Committee. Face validity of the measure score as an indicator of quality was systematically assessed as follows: A review of relevant evidence and guidelines and expert panel consensus process resulted in determined that this is a valid measure of quality of stroke care for patients receiving an IV thrombolytic therapy within 60 minutes. The expert panel was asked to rate their agreement with the following statement: "The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality

Testing results:

Prior testing results:

Empirical testing results:

- This measure had a inter-rater reliability of 0.72 (kappa statistic, 95% confidence interval 0.37-1.00.) including reliability of Door to IV tPA <= 60 minutes (yes/no.) . According to the Landis & Koch classification, a kappa of 0.72 is interpreted as substantial agreement.

Face Validity testing results:

- 100% of the expert panel strongly agreed with the value of this measure.

Updated results:

Face validity testing results :

- 85% of respondents either agreed or strongly agreed that this measure can accurately distinguish good and poor quality

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

☒ **Face validity only**

☒ **Empirical validity testing of the measure score**

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:
Exclusions include:

- Age < 18 years;
- Stroke occurred after hospital arrival (in ED/Obs/inpatient);
- Patients whose date/time of ED arrival and/or date/time of thrombolytic administration are blank, not documented, or N/A;
- Patients with a negative calculated time difference;
- Patients with a Date Last Known Well, but no time Last Known Well just MM/DD/YYYY;
- Patients that receive tPA greater than 4.5 hours after Last Known Well Patients transferred from outside hospital;
- Clinical Trial.

Exceptions include:

- Documented eligibility or medical reason for delay in treatment.

Results of the exclusions analysis:

- 672 hospitals with the minimum (10) number of quality reporting events, had 1,950 exclusions
- Average number of exclusions per hospital is 2.90.
- Overall exclusion rate is 10.8%.
- Range of exclusion rates for hospitals is 47% to 0%.

Exceptions Analysis:

- 672 hospitals with the minimum (10) number of quality reporting events had 3,581 exceptions.
- Average number of exceptions per hospital is 5.32.
- Overall exception rate is 18.2%.
- Range of exclusion rates for hospitals included 57% to 0%.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*
- *Do the number and percent of exceptions raise any concerns?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

Measures of central tendency, variability, and dispersion were calculated.

Hospital	Mean Rate	Median Rate	Mode	SD	Range of Rate	Minimum	Maximum
672	0.70	0.73	1.0	0.22	1.0	0.0	1.0

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- Data is collected through a clinical registry using the Get with the Guidelines Stroke Data Collection Form. This is a paper version of the electronic data collection tool which is called the Patient Management Tool (PMT). This is the only way

2b7. Missing Data

- The developer describes the process and criteria that is relevant for this measure. If data required to determine patient eligibility are missing, those patients/cases would be ineligible for inclusion in the denominator and therefore the patient/case would be deleted.
- If the data required to determine if a denominator eligible patient qualifies for the numerator (or has a valid exclusion/exception) are missing, this case would represent a quality failure and would not be included.

Questions for the Committee:

- *Are there any concerns regarding how missing data are treated?*

Guidance from the Validity Algorithm: specifications consistent with evidence (Box 1) → threats mostly assessed (Box2) → empirical testing conducted (box 3) → data element level testing (Box 10) → method appropriate (box 11) → moderate certainty that data are valid (Box 12a)

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **This measure is valid but does not capture full time from stroke onset to treatment which is the core measure for studies of IV TPA in acute ischemic stroke
**specifications consistent with the evidence

2a2. Reliability Testing

Comments: **I could find no validity testing for this measure
**The measure addresses the evidence effectively however it extends the treatment beyond the evidence by not addressing the issue of mobility. Several RCTs and reviews support leave out the mobile stroke patient from studies of VTE prophylaxis.
**new validity testing demonstrates sufficient validity

2b2. Validity Testing

Comments: **Measure excludes stroke patients under 18 years of age. However all data generated on populations 18 or older so this is consistent with the studied populations.
Using median time to admin of TPA in a population may be affected by skewed deviations (outliers) affecting performance.
**Documentation of why no VTE prophylaxis is considered a positive response and is a quite permissive criterion.
**No

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **I could find not reliability testing for this measure

**The measure appears to be reliable, with good results for inter-rater reliability with a large N (77 hospitals, 739 patients). Overall agreement rate = 98%. Reliability was tested at the data element level.
**No need for new reliability testing

Criterion 3. Feasibility
Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- This data is collected through a clinical registry, the Get With the Guidelines – Stroke
- The developer states that there are no issues with data collection have been identified and no modifications have been made to this measure, as collected in the GWTG – Stroke registry, due to issues with data collection, sampling or cost.
- The data for this measures is abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments: **This is one of the routinely measured components of acute stroke care. THis is already systematically measures in health systems (Stroke Centers) and reported.
**The measure is quite feasible as evidenced by prior adoption.
**Multiple ways that data collection is implemented- no concerns.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

- Professional Certification or Recognition Program -Stroke Hospital Recognition Program through the Get with the Guidelines-Stroke; Details of the program including a Web-based Patient Management Tool™, decision support, a robust registry, real-time benchmarking capabilities are provided.
- Quality Improvement with Benchmarking -Stroke Hospital Recognition Program; Achievement Awards recognize hospitals that demonstrate at least 85 percent compliance in each of 7 Get With The Guidelines-stroke Achievement Measures.
- Quality Improvement (Internal to the specific organization); participating hospitals commit to reaching the Target: Stroke performance goal of 50 percent or more of eligible patients treated with thrombolytic within 60 minutes of hospital arrival.
- NQF includes the use of performance results about identifiable accountable entities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**
OR
Planned use in an accountability program? ☐ **Yes** ☒ **No**

Accountability program details

- Professional Certification or Recognition Program -Stroke Hospital Recognition Program through the Get with the Guidelines-Stroke; Details of the program including a Web-based Patient Management Tool™, decision support, a robust registry, real-time benchmarking capabilities
- Quality Improvement with Benchmarking -Stroke Hospital Recognition Program; Achievement Awards recognize hospitals that demonstrate at least 85 percent compliance in each of 7 Get With The Guidelines-stroke Achievement Measures

Improvement results

- The developer presented data demonstrating a mean performance score from the Get with the Guidelines registry that increased from 53% to 70% (2012-2015). This indicates that the implementation of the measure led to a positive trend in the proportion of patients receiving tPA treatment with a door-to-needle time of 60 minutes or less.

Unexpected findings (positive or negative) during implementation
None stated

Potential harm and unintended consequences

- The developer reports that numerous studies has shown that door-to- needles times has not led to unintended consequences such as increase in mortality or increase in bleeding complications.
- Studies looked at a reduced use of tPA, which could happen if hospitals were neglecting to treat after 60 minutes to avoid a door to needle time <=60 minutes failure on the measure as well as others.
- Developer highlights that the rate of symptomatic intracerebral hemorrhage in tPA-treated patients with door-to-needle times within 60 minutes are lower than the rate of those who did not receive treatment within 60 minutes of arrival at the hospital. The rate of those who received tPA within 60 minutes was 4.7%, lower than the 5.6% rate of reported in patients with door-to-needle times >60 minutes.

Feedback: There is no feedback on this measure.
***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **http://www.jointcommission.org/stroke/
**The measure is usable as evidenced by prior adoption, in CY2014 it was adopted at 1,299 hospitals for 213,000 cases. This is 23%

of 5627 hospitals in the US.
**Accountability is high- public reporting and Hospital IQR

Criterion 5: Related and Competing Measures

Competing measures:

- The current measure (#1952) captures acute ischemic stroke patients aged 18 years and older receiving intravenous tissue plasminogen activator (tPA) therapy during the hospital stay and having a time from hospital arrival to initiation of thrombolytic therapy administration (door-to-needle time) of 60 minutes or less.
- # 0437 STK 04: Thrombolytic Therapy
  This measure captures the proportion of acute ischemic stroke patients who arrive at this hospital within 2 hours of time last known well for whom IV t-PA was initiated at this hospital within 3 hours of time last known well

Harmonization:

These measures are not harmonized.

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

1NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 1952

Measure Title: Time to Intravenous Thrombolytic Therapy

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 1/15/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note:

A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

- ☐ Health outcome: Click here to name the health outcome
- ☐Patient-reported outcome (PRO): Click here to name the PRO

  PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

- ☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☒ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

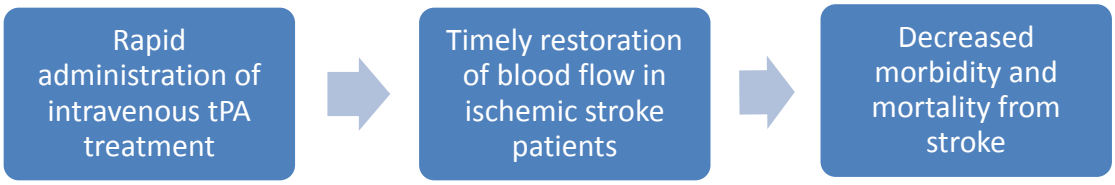
1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Jauch EC, Saver JL, Adams HP, Bruno A, Connors JJ, Demaerschalk BM, et al; American Heart Association Stroke Council; Council on Cardiovascular Nursing; Council on Peripheral Vascular Disease; Council on Clinical Cardiology. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44(3):870-947.

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

AHA/ASA 2013 Guideline:

In patients eligible for intravenous rtPA, benefit of therapy is time dependent, and treatment should be initiated as quickly as possible. The door-to-needle time (time of bolus administration) should be within 60 minutes from hospital arrival. (Class I; Level of Evidence A) p. 898

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

AHA/ASA 2013 Guideline:

The AHA/ASA recommendation included in section 1a.4.2. has been assigned a Class I. Class I recommendations refer to "Conditions for which there is evidence for and/or general agreement that the procedure or treatment is useful and effective."

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

AHA/ASA 2013 Guidelines:

The standard AHA algorithm for classifying recommendations and levels of evidence

focuses on therapeutic questions and, consequently, emphasizes evidence from randomized clinical

trials. As such, AHA/ASA guideline methodology categorizes indications as class I, II, or III on the basis of a multifactorial assessment of risk and expected efficacy viewed in the context of current knowledge and the relative strength of this knowledge. These classes summarize the recommendations for procedures or treatments as follows and noted in the table below:

Class I: Conditions for which there is evidence and/or general agreement that a given procedure or treatment is useful and effective.

Class II: Conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment.

- IIa: Weight of evidence/opinion is in favor of usefulness/efficacy
- IIb: Usefulness/efficacy is less well established by evidence/opinion.

Class III: Conditions for which there is evidence and/or general agreement that the procedure/treatment is not useful/effective e and in some cases may be harmful.

- No Benefit- Procedure/Test not helpful or Treatment w/o established proven benefit
- Harm- Procedure/Test leads to excess cost w/o benefit or is harmful, and or Treatment is harmful

Additional detail regarding the classification of recommendation and level of evidence is provided in the following table.

Table 1. Applying Classification of Recommendations and Level of Evidence

ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT	SIZE OF TREATMENT EFFECT			
	CLASS I *Benefit >>> Risk* Procedure/Treatment **SHOULD** be performed/ administered	**CLASS IIa** *Benefit >> Risk* *Additional studies with focused objectives needed* **IT IS REASONABLE** to perform procedure/administer treatment	**CLASS IIb** *Benefit ≥ Risk* *Additional studies with broad objectives needed; additional registry data would be helpful* Procedure/Treatment **MAY BE CONSIDERED**	**CLASS III** *No Benefit* or **CLASS III** *Harm* Procedure/ Test / Treatment COR III: No benefit / Not Helpful / No Proven Benefit COR III: Harm / Excess Cost w/o Benefit or Harmful / Harmful to Patients
LEVEL A Multiple populations evaluated* Data derived from multiple randomized clinical trials or meta-analyses	■ Recommendation that procedure or treatment is useful/effective ■ Sufficient evidence from multiple randomized trials or meta-analyses	■ Recommendation in favor of treatment or procedure being useful/effective ■ Some conflicting evidence from multiple randomized trials or meta-analyses	■ Recommendation's usefulness/efficacy less well established ■ Greater conflicting evidence from multiple randomized trials or meta-analyses	■ Recommendation that procedure or treatment is not useful/effective and may be harmful ■ Sufficient evidence from multiple randomized trials or meta-analyses
LEVEL B Limited populations evaluated* Data derived from a single randomized trial or nonrandomized studies	■ Recommendation that procedure or treatment is useful/effective ■ Evidence from single randomized trial or nonrandomized studies	■ Recommendation in favor of treatment or procedure being useful/effective ■ Some conflicting evidence from single randomized trial or nonrandomized studies	■ Recommendation's usefulness/efficacy less well established ■ Greater conflicting evidence from single randomized trial or nonrandomized studies	■ Recommendation that procedure or treatment is not useful/effective and may be harmful ■ Evidence from single randomized trial or nonrandomized studies
LEVEL C Very limited populations evaluated* Only consensus opinion of experts, case studies, or standard of care	■ Recommendation that procedure or treatment is useful/effective ■ Only expert opinion, case studies, or standard of care	■ Recommendation in favor of treatment or procedure being useful/effective ■ Only diverging expert opinion, case studies, or standard of care	■ Recommendation's usefulness/efficacy less well established ■ Only diverging expert opinion, case studies, or standard of care	■ Recommendation that procedure or treatment is not useful/effective and may be harmful ■ Only expert opinion, case studies, or standard of care
Suggested phrases for writing recommendations	should is recommended is indicated is useful/effective/beneficial	is reasonable can be useful/effective/beneficial is probably recommended or indicated	may/might be considered may/might be reasonable usefulness/effectiveness is unknown/unclear/uncertain or not well established	COR III: No Benefit: is not recommended; is not indicated; should not be performed/ administered/ other; is not useful/ beneficial/ effective COR III: Harm: potentially harmful; causes harm; associated with excess morbidity/mortality; should not be performed/ administered/ other
Comparative effectiveness phrases†	treatment/strategy A is recommended/indicated in preference to treatment B treatment A should be chosen over treatment B	treatment/strategy A is probably recommended/indicated in preference to treatment B it is reasonable to choose treatment A over treatment B		

A recommendation with Level of Evidence B or C does not imply that the recommendation is weak. Many important clinical questions addressed in the guidelines do not lend themselves to clinical trials. Although randomized trials are unavailable, there may be a very clear clinical consensus that a particular test or therapy is useful or effective.

*Data available from clinical trials or registries about the usefulness/efficacy in different subpopulations, such as sex, age, history of diabetes, history of prior myocardial infarction, history of heart failure, and prior aspirin use.

†For comparative effectiveness recommendations (Class I and IIa; Level of Evidence A and B only), studies that support the use of comparator verbs should involve direct comparisons of the treatments or strategies being evaluated.

Additional detail regarding AHA/ASA's gradation recommendations is provided in the following table.

Table 2. Definition of Classes and Levels of Evidence Used in AHA/ASA Recommendations

Class I	Conditions for which there is evidence for and/or general agreement that the procedure or treatment is useful and effective.
Class II	Conditions for which there is conflicting evidence and/or a divergence of opinion about the usefulness/efficacy of a procedure or treatment.
Class IIa	The weight of evidence or opinion is in favor of the procedure or treatment.
Class IIb	Usefulness/efficacy is less well established by evidence or opinion.
Class III	Conditions for which there is evidence and/or general agreement that the procedure or treatment is not useful/effective and in some cases may be harmful.
Therapeutic recommendations	
Level of Evidence A	Data derived from multiple randomized clinical trials or meta-analyses
Level of Evidence B	Data derived from a single randomized trial or nonrandomized studies
Level of Evidence C	Consensus opinion of experts, case studies, or standard of care
Diagnostic recommendations	
Level of Evidence A	Data derived from multiple prospective cohort studies using a reference standard applied by a masked evaluator
Level of Evidence B	Data derived from a single grade A study or 1 or more case-control studies, or studies using a reference standard applied by an unmasked evaluator
Level of Evidence C	Consensus opinion of experts

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

AHA/ASA 2013 Guidelines:

ACCF/AHA Task Force on Practice Guidelines. Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines. American College of Cardiology Foundation and American Heart Association, Inc. Cardiosource.com. 2010. Available at: http://assets.cardiosource.com/Methodology_Manual_for_ACC_AHA_Writing_Committees.pdf and http://my.americanheart.org/idc/groups/ahamah-public/@wcm/@sop/documents/downloadable/ucm_319826.pdf.

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

AHA/ASA 2013 Guidelines:

The section of the AHA/ASA guideline which includes the recommendations referenced in 1a4.2. pertains to the efficacy of intravenous rtPA and benefit of rapid initiation to improve patient outcomes after an acute ischemic stroke.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

AHA/ASA 2013 Guidelines:

The weight of the evidence in support of the listed AHA/ASA recommendations included in section 1a.4.2 is rated as Level A, as noted parenthetically. Level A evidence refers to “Data derived from multiple randomized clinical trials or meta-analyses.”

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

AHA/ASA 2013 Guidelines:

Levels A evidence is described in 1a.7.2. Level B evidence refers to "Data derived from a single randomized trial, or nonrandomized studies." Level C evidence refers to "Only consensus opinion of experts, case studies, or standard-of-care." Additional details and information about the evidence rating scheme can also be seen in 1a.4.2. and 1a.4.3.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*).
Date range: 1995-2012

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

AHA/ASA 2013 Guidelines:

Information regarding the total number of studies and type of study designs included in the body of evidence is not available.

However, the guidelines cite that 16 randomized control trials, 1 open trial, 32 observational studies, and 4 meta-analyses were reviewed to develop the recommendations provided in 1a4.2 and most relevant to the patient populations addressed in the measure.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

AHA/ASA 2013 Guidelines:

Information regarding the overall quality of evidence across studies is not available.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

AHA/ASA 2013 Guidelines:

The guideline does not include an overall estimate of benefit from the body of evidence. However, they do include the following summary information regarding the benefits of timely rtPA treatment, "Intravenous administration of rtPA remains the only FDA-approved pharmacological therapy for treatment of patients with acute ischemic stroke. Its use is associated with improved outcomes for a broad spectrum of patients who can be treated within 3 hours of the last known well time before symptom onset and a mildly more selective spectrum of patients who can be treated between 3 and 4.5 hours of the last known well time. Most importantly, earlier treatment is more likely to result in a favorable outcome."

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

Harms studied focused on the harms of treatment rather than the harms of time-initiated therapy.

As noted in the AHA/ASA guidelines, intracranial hemorrhages were reported in community based-settings prior to the approval of rtPA as a treatment option. However, the guidelines state it is now clear that the risk of hemorrhage is proportional to the degree to which the NINDS protocol is not followed. Other adverse events studied include systemic bleeding, myocardial rupture if fibroinolytics are given within a few days of acute myocardial infarction and reactions such as anaphylaxis or angioedema also has occurred, but these events are rare. Orolingual angioedema reactions have occurred in 1.3%-5.1% of patients, however, reactions are typically mild. Despite the harms listed, it is ultimately determined that the benefits of timely treatment outweigh all harms studied.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

1. Citation: Fonarow GC, Zhao X, Smith EE, Saver JL, Reeves MJ, Bhatt DL, et al. Door-to-needle times for tissue plasminogen activator administration and clinical outcomes in acute ischemic stroke before and after a quality improvement initiative. JAMA. 2014;311(16):1632-1640. DOI:10.1001/jama.2014.3203.

   Description: Observational study as a part of Target Stroke Initiative that looked to evaluate door-to-needle times for tPA administration. Additionally the study evaluated the proportion of patients with door-to-needle times of ≤ 60 minutes before and after a quality improvement initiative to determine if improvements in door-to-needle times were associated with improved clinical outcomes.

   Results: "Importantly, the improvement in timeliness in tPA administration following the start of the program was associated with improved clinical outcomes including lower in-hospital mortality, more frequent discharge to a more independently functioning environment, and lower rates of tPA complications, including symptomatic intracranial hemorrhage. These findings further reinforce the importance and clinical benefits of more rapid administration of intravenous tPA."

   Conclusion: This study further highlights the importance of a door-to-needle time of ≤ 60 minutes for the administration of tPA following an ischemic stroke. While timely administration leads to improved

clinical outcomes, the study highlights that less than 30% of patients are receiving treatment within the recommended timeline and further emphasizes the opportunity for improvement for facilities.

2. Citation: Emberson J, Lees KR, Lyden P, Blackwell L, Albers G, Bluhmki E, et al. Effect of treatment delay, age, and stroke severity on the effects of intravenous thrombolysis with alteplase for acute ischaemic stroke: a meta-analysis of individual patient data from randomized trials. Lancet. 2014;384(9958):1929-1935.

   Description: Meta-analysis of individual patient data from 6756 patients in nine randomized trials comparing alteplase with placebo or open control. The primary goal of the analysis was to explore the extent to which treatment delay affected the effect of the alteplase and to establish if age or stroke severity affected treatment effects. The authors defined good stroke outcome as no significant disability at 3-6 months as defined by a modified Rankin scale of 0 or 1. Additional outcomes included symptomatic intracranial hemorrhage, fatal intracranial hemorrhage within 7 days and 90-day mortality.

   Results: “Alteplase significantly increased the odds of a good outcome, with earlier treatment resulting in significantly greater proportional benefit increasing proportional benefit with earlier treatment.” The study also states, “The effect of alteplase on a good outcome was chiefly driven by treatment delay; after controlling for treatment delay, neither age nor severity of stroke contributed significant additional predictive value.”

   The tables below demonstrate the importance of early treatment for improved outcomes.

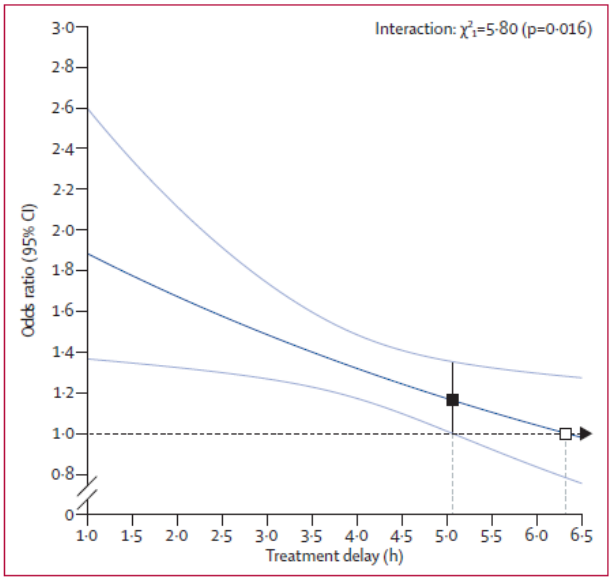


Figure 1: **Effect of timing of alteplase treatment on good stroke outcome (mRS 0–1)**
The solid line is the best linear fit between the log odds ratio for a good stroke outcome for patients given alteplase compared with those given control (vertical axis) and treatment delay (horizontal axis; $p_{interaction}$=0·016). Estimates are derived from a regression model in which alteplase, time to treatment, age, and stroke severity (handled in a quadratic manner) are included as main effects but the only treatment interaction included is with time to treatment. Only 198 patients (159 from IST–3) had a time from stroke onset to treatment of more than 6 h. The white box shows the point at which the estimated treatment effect crosses 1. The black box shows the point at which the lower 95% CI for the estimated treatment effect first crosses 1·0. mRS=modified Rankin Scale.

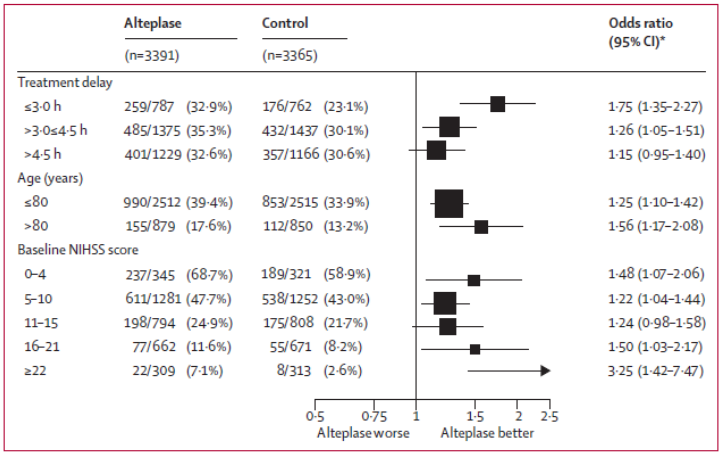


Figure 2: **Effect of alteplase on good stroke outcome (mRS 0–1), by treatment delay, age, and stroke severity**
*For each of the three baseline characteristics, estimates were derived from a single logistic regression model stratified by trial, which enables separate estimation of the OR for each subgroup after adjustment for the other two baseline characteristics (but not for possible interactions with those characteristics). mRS=modified Rankin Scale.

Conclusion: The results of the meta-analysis indicate that the early administration of tPA from symptom onset is effective in improving good outcomes for stroke patients. The meta-analysis further emphasizes the importance of timely administration of tPA and shows a proportional benefit with earlier treatment.

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
1952_evidence_attachment-635883916612584511.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Multiple studies have shown that the rapid administration of intravenous recombinant tissue-type plasminogen activator (tPA) to appropriate patients is a proven, effective treatment in restoring blood flow and reducing long-term morbidity for ischemic stroke patients.

With every 10-minute delay in the start of thrombolytic infusion within the 1-to-3 hour-treatment time period, there was 1 fewer patient of 100 patients having improved disability outcome. The administration of tPA is time-dependent and shortening door-to-needle time helps improve functional outcomes at 4 to 6 months when given in the recommended timeframe (8, 9). The Target Stroke Initiative to improve the timeliness of tPA administration found that among 71,169 patients with acute ischemic stroke that were treated with tPA at 1,030 GWTG stroke participating hospitals, participation in the program was associated with decreased door-to-needle time, lower in-hospital mortality and intracranial hemorrhage, and an increase in the percentage of patients discharged home (10). Results from clinical trials and registries have encouraged multiple organizations to set targets for timely initiation of thrombolytic therapy after hospital arrival. A National Institute of Neurological Disorders and Stroke national symposium on the rapid identification and treatment of acute stroke recommended a door-to-needle target time of 60 minutes (1). American Heart Association/American Stroke Association guidelines recommend the target for completion of initial evaluation and start of tPA treatment should be within 60 minutes of the patient´s arrival in the emergency department (2-3). The Brain Attack Coalition´s target for primary stroke centers is to achieve a door-to-needle time within 60 minutes in 80% or more of patients (4). Despite evidence and recommendations advocating a door-to-needle time of 60 minutes or less, studies have shown that less than 30% of US patients are treated within this time window (5-7). This measure is intended to promote a reduction in door to needle times and improvement in the proportion of eligible patients receiving treatment within 60 minutes of hospital arrival.

Citations:

1. Marler JR, Winters Jones P, Emr M. The National Institute of Neurological Disorders and Stroke: Proceedings of National Symposium on Rapid Identification and Treatment of Acute Stroke. Bethesda, MD: National Institute of Neurological Disorders and Stroke, 1997.

2. Adams HP Jr., del Zoppo G, Alberts MJ, Bhatt DL, Brass L, Furlan A, et al. Guidelines for the early management of adults with Ischemic stroke. Circulation. 2007;115:e478–534. Available at: http://circ.ahajournals.org/content/115/20/e478.full

3. Summers D, Leonard A, Wentworth D, Saver JL, Simpson J, Spilker JA, et al. Comprehensive overview of nursing and interdisciplinary care of the acute ischemic stroke patient: a scientific statement from the American Heart Association. Stroke. 2009;40:2911–2944. Available at: http://stroke.ahajournals.org/content/40/8/2911.full.pdf+html

4. Alberts MJ, Hademenos G, Latchaw RE, Jagoda A, Marler JR, Mayberg MR, et al. Recommendations for the establishment of primary stroke centers. Brain Attack Coalition. JAMA. 2000;283:3102–3109. Available at: http://jama.ama-assn.org/content/283/23/3102.full

5. Fonarow GC, Zhao X, Smith EE, Saver JL, Reeves MJ, Bhatt DL, et al. Door-to-needle times for tissue plasminogen activator administration and clinical outcomes in acute ischemic stroke before and after a quality improvement initiative. JAMA. 2014;311(16):1632-1640. DOI:10.1001/jama.2014.3203.

6. Jauch EC, Saver JL, Adams HP, Bruno A, Connors JJ, Demaerschalk, BM, et al; American Heart Association Stroke Council; Council on Cardiovascular Nursing; Council on Peripheral Vascular Disease; Council on Clinical Cardiology. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44(3):870-947.

7. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics- 2016 update a report from the American Heart Association. Circulation. 2015;132:000-000. DOI: 10.1161/CIR.0000000000000350.

8. Lansberg MG, Schrooten M, Bluhmki E, Thijs VN, Saver JL. Treatment time-specific number needed to treat estimates for tissue plasminogen activator therapy in acute stroke based on shifts over the entire range of the modified Rankin Scale. Stroke. 2009;40:2079 –2084. Available at: http://stroke.ahajournals.org/content/40/6/2079.full

9. Powers WJ, Derdeyn CP, Biller J, Coffey CS, Hoh BL, Jauch EC, et al. 2015 American Heart Association/American Stroke Association focused update of the 2013 guidelines for the early management of patients with acute ischemic stroke regarding endovascular treatment. Stroke. 2015;46: 3024-3039. DOI:10.1161/STR.0000000000000074

10. Fonarow G, Smith EE, Saver J, Reeves M, Bhatt D, Grau-Sepulveda M, Olson D, Hernandez A, Peterson E, Schwamm L. Timeliness of Tissue-Type Plasminogen Activator Therapy in Acute Ischemic Stroke: Patient Characteristics, Hospital Factors, and Outcomes Associated With Door-to-Needle Times Within 60 Minutes. Circulation. 2011, 123:750-758: originally published online February 10, 2011 doi: 10.1161/CIRCULATIONAHA.110.974675. Available at: Available at: http://circ.ahajournals.org/content/123/7/750.full.pdf

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Get With The Guidelines – Stroke (GWTG-Stroke) is a clinical data registry that collects information from hospitals and clinicians on patient demographics, acute outcomes, quality measures, and health outcomes. The registry was piloted in 2001 and nationally implemented in 2003. GWTG-Stroke currently has 1,656 hospitals participating in the program and is managed by the American Heart Association (AHA) and American Stroke Association (ASA).

10/1/2014 – 9/30/2015 Performance Data:
Mean: 70%
Standard Error: .01
Median: .73
Standard Deviation: .22
Minimum: 0.00
Maximum: 1.00

Interquartile Range:	Result %
25	56.00%
50	73%
75	88%
100	100%

Decile	Result
1	40.00%
2	50.00%
3	60.00%
4	68.00%
5	73.00%
6	79.00%
7	84.00%
8	90.00%
9	95.00%
10	100.00%

01/01/2012 – 09/30/2014 Performance Data

Mean: 53.00%
Standard Error: .01
Median: .55
Standard Deviation: 0.21
Minimum: 0.00
Maximum: 0.95

Interquartile Range	Result
25	38.00%
50	55.00%
75	70.00%
100	95.00%

Decile	Result
1	0.23
2	0.33
3	0.42
4	0.49
5	0.55
6	0.61
7	0.67
8	0.72
9	0.79
10	0.95

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

Recent studies have indicated that only 30% of patients receive tPA treatment within the guideline-recommended 60 minute door-to-needle time (1). A study conducted by Fonarow and colleagues conducted an analysis of tPA-treated patients in the GWTG-Stroke program from 2003-2009 and found and found that among 25,504 ischemic stroke patients from 1,082 hospitals treated with tPA, door-to-needle time was <=60 minutes in only 6,790 (26.6%). This number only increased moderately from 19% in 2003 to 29% in 2009 (2). The study also found that there is substantial variation between hospitals in the proportion of ischemic stroke patients with door-to needle times within 60 minutes. Of the 1,082 US hospitals studied, the hospital level rates of treatment vary widely with a median rate of 21.1% (25th to 75th percentile, 13.0% to 33.3%) and range of 0% to 79.2%. Among hospitals with at least 10 patients treated with tPA within 3 hours of symptom onset, the proportion of patients in whom door-to-needle times within 60 minutes were achieved was 0% to <=20% at 290 hospitals (45.2%), 21% to <=40% at 242 (37.8%), 40% to <=60% at 95 (14.8%), and 60% to <=80% at 14 (2.2%). This study concluded that fewer than one-third of patients treated with intravenous tPA had door-to-needle times <=60 minutes, with only modest improvement over 6.5 years. These findings support the need for an NQF endorsed performance measure along with targeted initiatives to improve the timeliness of reperfusion in acute ischemic stroke.

Citations:

1. Fonarow GC, Zhao X, Smith EE, Saver JL, Reeves MJ, Bhatt DL, et al. Door-to-needle times for tissue plasminogen activator administration and clinical outcomes in acute ischemic stroke before and after a quality improvement initiative. JAMA. 2014;311(16):1632-1640. DOI:10.1001/jama.2014.3203.

2. Fonarow G, Smith EE, Saver J, Reeves M, Bhatt D, Grau-Sepulveda M, Olson D, Hernandez A, Peterson E, Schwamm L. Timeliness of Tissue-Type Plasminogen Activator Therapy in Acute Ischemic Stroke: Patient Characteristics, Hospital Factors, and Outcomes Associated With Door-to-Needle Times Within 60 Minutes. Circulation. 2011, 123:750-758: originally published online February 10, 2011 doi: 10.1161/CIRCULATIONAHA.110.974675. Available at: Available at: http://circ.ahajournals.org/content/123/7/750.full.pdf

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

Get With The Guidelines – Stroke (GWTG-Stroke) is a clinical data registry that collects information from hospitals and clinicians on patient demographics, acute outcomes, quality measures, and health outcomes. The registry was piloted in 2001 and nationally implemented in 2003. GWTG-Stroke currently has 1,656 hospitals participating in the program and is managed by the American Heart Association (AHA) and American Stroke Association (ASA).

10/1/2014 – 9/30/2015 Performance Data Across Different Demographic Variables:Age < 65: 73.64%

Ages 65-79: 74.25%

Age >80: 72.84%

Male: 75.46%

Female: 71.65%

Hispanic: 72.06%

Black: 74.10%

Indian: 72.97%

Asian: 77.64%

Hawaiian: 80.48%

01/01/2012 – 09/30/2014 Performance Data Across Different Demographic Variables:

Age < 65: 58.49%

Ages 65-79: 59.83%

Age >80: 58.69%

Male: 60.55%

Female: 57.41%

Hispanic: 60.77%

Black: 58.05%

Indian: 55.65%

Asian: 62.45%

Hawaiian: 68.07%

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

A study conducted by Fonarow and colleagues examined data from acute ischemic stroke patients treated with tPA within 3 hours of symptom onset to determine patient and hospital characteristics, frequency and temporal trends of patients that were administered tPA within 60 minutes of arrival time. Data was gathered from 1082 hospitals participating in the GWTG-Stroke program from April 1, 2003 to September 30, 2009. Fonarow found that older patients, black patients, and those with less severe strokes or arriving during off hours were less likely to receive timely care. The study also states that patients who were more strongly associated to receive tPA administration within 60 minutes were younger, male gender, white, and had no history of a stroke. Additionally, there was variation in care provided by hospitals. Hospitals that had less experience providing tPA to ischemic stroke patients were less likely to administer the therapy within 60 minutes.

Citation:
Fonarow G, Smith EE, Saver J, Reeves M, Bhatt D, Grau-Sepulveda M, Olson D, Hernandez A, Peterson E, Schwamm L. Timeliness of Tissue-Type Plasminogen Activator Therapy in Acute Ischemic Stroke: Patient Characteristics, Hospital Factors, and Outcomes Associated With Door-to-Needle Times Within 60 Minutes. Circulation. 2011, 123:750-758: originally published online February 10, 2011 doi: 10.1161/CIRCULATIONAHA.110.974675. Available at: Available at: http://circ.ahajournals.org/content/123/7/750.full.pdf

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Stroke is considered the number 5 killer in the United States. On average, a stroke occurs every four minutes in the United States. It is estimated that strokes account for approximately one out of every twenty deaths in the United States and that 6.6 million Americans have suffered from a stroke. Each year, approximately 795,000 people experience a new or recurrent stroke. Overall stroke prevalence is an estimated 2.6%. Of all strokes, 87% are ischemic strokes. Stroke is a leading cause of long-term disability in the United States and was considered among the top 18 diseases contributing to years lived with disability in 2010. The 30-day hospital readmission rate after discharge from postacute rehabilitation for stroke is 12.7% among fee-for service Medicare patients. The mean rehabilitation length of stay for stroke is 14.6 days. In 2011-2012, the direct and indirect cost of stroke was $33.0 billion. The estimated direct medical cost of stroke was $17.2 billion, included hospital outpatient or office-based provider visits, hospital inpatient stays, ED visits, prescribed medicines, and home health care.

Administrative data found that between 3.45% and 5.2% of acute ischemic stroke patients were treated with tPA therapy in 2009. Similarly, an analysis of the GWTG-Stroke program demonstrated an increase in treatment rates from 2003 to 2011.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, et al. Heart disease and stroke statistics- 2016 update a report from the American Heart Association. Circulation. 2015;132:000-000. DOI: 10.1161/CIR.0000000000000350.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable. Not a PRO-PM.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology, Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
The measure specifications are included as an attachment with this submission. Additional measure details may be found at http://stroke.ahajournals.org/content/45/11/3472.full.pdf+html

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Time_to_Thrombolytic_Data_Dictionary.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Supporting guidelines and technical specifications included in the measure are reviewed on an annual basis. This annual review resulted in changes to update the denominator details and exclusions to preserve the original measure intent. Specifications have been updated to maintain alignment with The Joint Commission Stroke specifications manual ICD-10 coding, and to align measure component requirements in the details sections with the phrases found in the registry data dictionary to provide a clearer illustration of the requirements.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Acute ischemic stroke patients aged 18 years and older receiving intravenous tissue plasminogen activator (tPA) therapy during the hospital stay and having a time from hospital arrival to initiation of thrombolytic therapy administration (door-to-needle time) of 60 minutes or less.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Once for each hospital visit for ischemic stroke during the measurement period.

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
All denominator patients with the following:
['Date/time IV thrombolytic therapy initiated' minus 'Arrival Date/Time'] <= 60 minutes

**Data elements referenced align with information found in S.19 'Time to Intravenous Thrombolytic Therapy Specifications.docx' attachment.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All acute ischemic stroke patients who received intravenous thrombolytic therapy during the hospital stay.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
An ICD-9-CM/ICD-10 Principal Diagnosis Code for acute ischemic stroke:
Diagnosis for ischemic stroke ICD-9: 433.01, 433.10, 433.11, 433.21, 433.31, 433.81, 433.91, 434.00, 434.01, 434.11, 434.91, 436
Diagnosis for ischemic stroke ICD-10: ICD-10: I63.00, I63.011, I63.012, 163.019, I63.02, I63.031, I63.032, I63.039, I63.09, I63.10, I63.111, I63.112, I63.119, I63.12, I63.131, I63.132, I63.139, I63.19, I63.20, I63.211, I63.212, I63.219, I63.22, I63.231, I63.232, I63.239, I63.29, I63.30, I63.311, I63.312, I63.319, I63.321, I63.322, I63.329, I63.331, I63.332, I63.339, I63.341, I63.342, I63.349, I63.39, I63.40, I63.411, I63.412, I63.419, I63.421, I63.422, I63.429, I63.431, I63.432, I63.439, I63.441, I63.442, I63.449, I63.49, I63.50, I63.511, I63.512, I63.519, I63.521, I63.522, I63.529, I63.531, I63.532, I63.539, I63.541, I63.542, I63.549, I63.59, I63.6, I63.8, I63.9
OR:
'Final clinical diagnosis related to stroke' = Ischemic Stroke
AND:
'IV tPA initiated at this hospital' = YES*

*Thrombolytic therapy for stroke includes: Activase, Alteplase, IV t-PA, or Recombinant t-PA Tissue plasminogen activator.
**Data elements referenced align with information found in S.19 'Time to Intravenous Thrombolytic Therapy Specifications.docx' attachment.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Denominator Exclusions:
•Patients less than 18 years of age
•Patient stroke occurred while in hospital
•Patients received in transfer from the inpatient, or outpatient of another facility
•Patients that receive tPA greater than 4.5 hours after Last Known Well
•Clinical trial

Denominator Exceptions:
Patients with documented Eligibility or Medical reason for delay in treatment [eg, social, religious, initial refusal, hypertension requiring aggressive control with intravenous medications, inability to confirm patients eligibility, or further diagnostic evaluation to confirm stroke for patients with hypoglycemia (blood glucose < 50); seizures, or major metabolic disorders, or management of concomitant emergent/acute conditions such as cardiopulmonary arrest, respiratory failure requiring intubation), or investigational or experimental protocol for thrombolysis.]

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The AHA/ASA distinguishes between measure exceptions and measure exclusions. Exclusions arise when the intervention required by the numerator is not appropriate for a group of patients who are otherwise included in the initial patient or eligible population of a measure (ie, the denominator). Exclusions are absolute and are to be removed from the denominator of a measure and therefore clinical judgment does not enter the decision. For measure 1952, Time to Intravenous Thrombolytic Therapy, exclusions include patients who are less than 18 years of age, patients whose stroke occurred while in the hospital, patients that received in transfer from the inpatient or outpatient of another facility, patients that receive tPA greater than 4.5 hours after Last Known Well, and patients enrolled in clinical trials. Exclusions are included in the measure specifications.

Measure Exceptions
Exceptions are used to remove a patient from the denominator of a performance measure when the patient does not receive a therapy or service AND that therapy or service would not be appropriate due to patient-specific reasons. The patient would otherwise meet the denominator criteria. Exceptions are not absolute, and are based on clinical judgment, individual patient characteristics, or patient preferences. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception. Examples are provided in the measure exception language of instances that may constitute an exception and are intended to serve as a guide to clinicians. For measure 1952, Time to Intravenous Thrombolytic Therapy exceptions may include medical reason(s) [eg, hypertension requiring aggressive control with intravenous medications, inability to confirm patient eligibility, or further diagnostic evaluation needed to confirm stroke for patients with hypoglycemia (blood glucose <50); seizures, major metabolic disorders, or management of concomitant emergent/acute conditions such as cardiopulmonary arrest, respiratory failure requiring intubation], or investigational or experimental protocol for thrombolysis, or eligibility reason(s) (eg, social, religious, initial refusal). Although this methodology does not require the external reporting of more detailed exception data, the AHA/ASA recommends that facilities document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The AHA/ASA also advocates the systematic review and analysis of each facility's exceptions data to identify practice patterns and opportunities for quality improvement.

Additional details are as follows:

Age' < 18 years
OR
['Date/time IV thrombolytic therapy initiated' minus 'Date/time Last Known Well'] > 4.5 hours
OR
'Patient location when stroke symptoms discovered' = stroke occurred after hospital 'Arrival Date/Time'
OR
'How patient arrived at your hospital' = transfer from other hospital
OR
'Was patient enrolled in a clinical trial in which patients with the same condition as the measure set were being studied' = yes
OR
If any of the following is unknown, blank, or incomplete (aka, missing time): 'Arrival Date/Time', 'Date/time IV thrombolytic therapy initiated', 'Date/time Last Known Well'

Measure Exceptions:
['Date/time IV thrombolytic therapy initiated' minus 'Arrival Date/Time'] > 60 minutes
AND
Eligibility Reason OR Medical Reason = Present

**Data elements referenced align with information found in S.19 'Time to Intravenous Thrombolytic Therapy Specifications.docx' attachment. Measure Exclusions:
'Age' < 18 years
OR
['Date/time IV thrombolytic therapy initiated' minus 'Date/time Last Known Well'] > 4.5 hours
OR
'Patient location when stroke symptoms discovered' = stroke occurred after hospital 'Arrival Date/Time'
OR
'How patient arrived at your hospital' = transfer from other hospital
OR
'Was patient enrolled in a clinical trial in which patients with the same condition as the measure set were being studied' = yes
OR
If any of the following is unknown, blank, or incomplete (aka, missing time): 'Arrival Date/Time', 'Date/time IV thrombolytic therapy initiated', 'Date/time Last Known Well'

Measure Exceptions:
['Date/time IV thrombolytic therapy initiated' minus 'Arrival Date/Time'] > 60 minutes
AND
Eligibility Reason OR Medical Reason = Present

**Data elements referenced align with information found in S.19 'Time to Intravenous Thrombolytic Therapy Specifications.docx' attachment.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Consistent with CMS' Measures Management System Blueprint and recent national recommendations put forth by the IOM and NQF to standardize the collection of race and ethnicity data, we encourage the results of this measure to be stratified by race, ethnicity, administrative sex, and payer.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
No risk adjustment or risk stratification

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
Rate is determined by calculating those eligible patients meeting the numerator specification divided by those meeting the denominator specification.
1) Check to see if there is an ICD-9/ICD-10 principal diagnosis of ischemic stroke; exclude those patients without an appropriate diagnosis code.
2) Check to see if patient had an inpatient stroke; exclude those patients with inpatient stroke
3) Check to see if patient is 18 years or older; exclude those patients less than 18 years of age
4) Check to see if patient is in a clinical trial; exclude those patients who were in a clinical trial
5) Check to see patient arrival date is documented; exclude those patients for which arrival date is unable to be determined (blank/unknown)
6) Check to see if patient arrival time is documented; exclude those patients for which arrival time is unable to be determined (blank/unknown)
7) Check to see if patient was transferred from another hospital; exclude those patients who were transferred from another hospital
8) Check to see if patient had IV thrombolytic therapy initiated; exclude those patients for whom IV thrombolytic therapy was not initiated
9) Check thrombolytic initiation date; exclude those patient for which thrombolytic initiation date is unable to be determined (blank/unknown)
10) Check thrombolytic initiation time; exclude those patients for which thrombolytic initiation time is unable to be determined (blank/unknown)
11) IV Thrombolytic Initiation Date/Time should not be less than (aka, should not be documented as occurring prior to) hospital

arrival date/time; exclude those patients for whom arrival IV thrombolytic initiation date/time is less than hospital arrival date/time
12) Check to see date/time last known well; exclude patients for whom date/time last known well is unable to be determined (blank/unknown)
13) Check to see timing in hours. Timing (IV Thrombolytic Initiation Date/Time - Date/Time Last Known well) should be less than or equal to 4.5 hours. If greater than 4.5 hours exclude patients.
14) If timing is less than or equal to 4.5 hours, check to see if timing for IV thrombolytic therapy time (IV Thrombolytic Initiation Date/Time - Arrival Date/Time) is less than or equal to 60 minutes. If time was greater than 60 minutes, determine if patient had a valid documented exception/reason for delay.

If the patient does not meet the numerator and a valid exception is not present, this case represents a quality failure.

For detailed measure algorithm see attached.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable. The measure is not based on a sample

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not applicable. The measure is not based on a survey.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Patient eligibility is determined by a set of defined criteria relevant to a particular measure. If data required to determine patient eligibility are missing, those patients/cases would be ineligible for inclusion in the denominator and therefore the patient/case would be deleted.

If data required to determine if a denominator eligible patient qualifies for the numerator (or has a valid exclusion/exception) are missing, this case would represent a quality failure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data : Registry

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Get with the Guidelines Stroke Data Collection Form. This is a paper version of the electronic data collection tool which is called the Patient Management Tool (PMT).

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable. The measure is not a composite.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF_1952_Time_to_IV_Thrombolytic_Therapy_Testing_Attachment.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 1952
Measure Title: Time to Intravenous Thrombolytic Therapy
Date of Submission: 1/15/2016
Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures

with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Get With The Guidelines – Stroke (GWTG-Stroke) is a clinical data registry that collects information from hospitals and clinicians on patient demographics, acute outcomes, quality measures, and health outcomes. The registry was piloted in 2001 and nationally implemented in 2003. GWTG-Stroke currently has 1,656 hospitals participating in the program and is managed by the American Heart Association (AHA) and American Stroke Association (ASA).

1.3. What are the dates of the data used in testing? 10/1/2014 – 9/30/2015

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan

☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

The total number of hospitals reporting on this measure is 841. Of those, 672 hospitals had all the required data elements and met the minimum number of quality reporting events (10) for inclusion in the reliability analysis. For this measure, 79.9 percent of hospitals are included in the analysis, and the average number of quality reporting events is 23.9 for a total of 16,100 events. The range of quality reporting events for 672 hospitals included is from 138 to 10. The average number of quality reporting events for the remaining 20.1 percent of hospitals who aren't included is 6.03.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

There were 16,100 patients included in this testing and analysis. These were the patients that were associated with hospitals who had 10 or more patients eligible for this measure.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

The same data sample was used for reliability testing and exclusions analysis.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Patient-level socio-demographic (SDS) variables were not captured as part of the testing.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Reliability of the computed measure score was measured as the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in hospital performance. Reliability at the level of the specific hospital is given by:

Reliability = Variance (hospital-to-hospital) / [Variance (hospital-to-hospital) + Variance (hospital-specific-error]

Reliability is the ratio of the hospital-to-hospital variance divided by the sum of the hospital-to-hospital variance plus the error variance specific to a hospital. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in hospital performance.

Reliability testing was performed by using a beta-binomial model. The beta-binomial model assumes the hospital performance score is a binomial random variable conditional on the hospital's true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates.

Reliability is estimated at two different points, at the minimum number of quality reporting events for the measure and at the mean number of quality reporting events per hospital.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

This measure has 0.63 reliability when evaluated at the minimum level of quality reporting events and 0.81 reliability at the average number of quality events.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Face validity of the measure score as an indicator of quality was systematically assessed as follows.

After the measure was fully specified, the expert panel was asked to rate their agreement with the following statement:

The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

Scale 1-5, where 1= Strongly Disagree; 3= Neither Agree nor Disagree; 5= Strongly Agree

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The expert panel included 20 members. Panel members were comprised of experts from the AHA Council on Stroke 2015-2016 Leadership Committee. The list of expert panel members is as follows:

Mat Reeves, BVSc, PhD, FAHA
Mai Nguyen-Huynh, MD, MAS
Judith Lichtman, PhD, MPH, FAHA
Edward Jauch, MD, MS, FAHA, FACEP
Jennifer Majersik, MD, MS
Kevin Sheth, MD, FAHA
Phillip Scott, MD
Walter N. Kernan, MD
Brett Cucchiara, MD, FAHA
Mary Ann Bauman, MD
Claranne Mathiesen, MSN, RN, CNRN, SCRN
Karen Furie, MD, MPH, FAHA
Salvador Cruz-Flores, MD, MPH, FAHA, FACP
Alejandro Rabinstein, MD, FAHA
Colin Derdeyn, MD, FAHA
N. Jennifer Klinedinst PhD, PH, RN, FAHA
Jose Romano, MD, FAHA, FAAN
Barbara Lutz, PhD, RN, CRRN, FAHA, FAAN
Pooja Khatri, MD, MSc, FAHA
Kyra Becker, MD

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The results of the expert panel rating of the validity statement were as follows: N = 20; Mean rating = 4.2 and 85% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality.

Frequency Distribution of Ratings

1 – 1 responses (Strongly Disagree)
2 – 0 responses
3 – 2 responses (Neither Agree nor Disagree)
4 – 8 responses
5 – 9 responses (Strongly Agree)

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Exclusions include:

- Age < 18 years;
- Stroke occurred after hospital arrival (in ED/Obs/inpatient);

- Patients whose date/time of ED arrival and/or date/time of thrombolytic administration are blank, not documented, or N/A;
- Patients with a negative calculated time difference;
- Patients with a Date Last Known Well, but no time Last Known Well just MM/DD/YYYY;
- Patients that receive tPA greater than 4.5 hours after Last Known Well Patients transferred from outside hospital;
- And Clinical Trial.

Exceptions include:

- Documented eligibility or medical reason for delay in treatment.

Exclusions and exceptions were analyzed for frequency across providers.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Exclusions Analysis:

Amongst the 672 hospitals with the minimum (10) number of quality reporting events, there were a total of 1,950 exclusions reported. The average number of exclusions per hospital in this sample is 2.90. The overall exclusion rate is 10.8%. The range of exclusion rates for hospitals included 47% to 0%.

Exceptions Analysis:

Amongst the 672 hospitals with the minimum (10) number of quality reporting events, there were a total of 3,581 exceptions reported. The average number of exceptions per hospital in this sample is 5.32. The overall exception rate is 18.2%. The range of exclusion rates for hospitals included 57% to 0%.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusions arise when patients who are included in the initial patient or eligible population for a measure do not meet the denominator criteria specific to the intervention required by the numerator. Exclusions are absolute and apply to all patients and, therefore, are not part of clinical judgment within a measure. Exclusions, including applicable value sets, are included in the measure specifications.

Exceptions are used to remove a patient from the denominator of a performance measure when the patient does not receive a therapy or service AND that therapy or service would not be appropriate due to patient-specific reasons.

Without these being removed, the performance rate would not accurately reflect the true performance of each facility, which would result in an increase in performance failures and false negatives.

AHA/ASA recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. AHA/ASA also advocates for the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Not applicable

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Not applicable

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable

2b4.9. Results of Risk Stratification Analysis:

Not applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Measures of central tendency, variability, and dispersion were calculated.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Based on the sample of 672 included hospitals, the mean performance rate is 0.70, the median performance rate is 0.73 and the mode is 1.0. The standard deviation is 0.22. The range of the performance rate is 1.0, with a minimum rate of 0.0 and a maximum rate of 1.00. The interquartile range is 0.32 (0.56 – 0.88).

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The range of performance from 0.00 to 1.00 suggests there's clinically meaningful variation across physicians' performance.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

This test was not performed for this measure.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

This test was not performed for this measure.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

This test was not performed for this measure.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Data are not available to complete this testing.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Data are not available to complete this testing.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Data are not available to complete this testing.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

There are clinical exclusion criteria that may not be part of standard electronic data sets found within the electronic medical records. The AHA/ASA has the ultimate goal to be able to extract all information electronically and plans to work to identify codes and/or value sets that would be needed to identify exclusions and to work with the appropriate organization(s) to develop and implement any additional codes needed to capture information such as medical reasons, patient reasons or system reasons to exclude a patient from the denominator.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Given that the data for this measure is collected through the Get With the Guidelines – Stroke registry, and is not collected in an electronic health record, no feasibility assessment was performed. No issues with data collection have been identified and no modifications have been made to this measure, as collected in the GWTG – Stroke registry, due to issues with data collection, sampling or cost.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

Not applicable.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
	Professional Certification or Recognition Program Get with the Guidelines -Stroke Hospital Recognition Program http://www.heart.org/HEARTORG/HealthcareResearch/GetWithTheGuidelines/GetWithTheGuidelines-Stroke/Recognition-from-Get-With-The-Guidelines-Stroke_UCM_308034_Article.jsp#.Vo6fo_LlvIU Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Get with the Guidelines-Stroke http://www.heart.org/HEARTORG/HealthcareResearch/GetWithTheGuidelines/Get-With-The-Guidelines-Stroke_UCM_306098_SubHomePage.jsp Quality Improvement (Internal to the specific organization) Get with the Guidelines-Stroke; Target: Stroke http://www.strokeassociation.org/STROKEORG/Professionals/TargetStroke/Target-Stroke_UCM_314495_SubHomePage.jsp

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

1) Name of program and sponsor: Get with the Guidelines-Stroke; Sponsor: American Heart Association/American Stroke Association

Purpose: Get With The Guidelines®(GWTG) - Stroke is an in-hospital program for improving stroke care by promoting consistent adherence to evidence-based treatment guidelines. Numerous published studies demonstrate the program´s success in achieving measurable patient outcome improvements. The program provides hospitals with a Web-based Patient Management Tool™, decision support, a robust registry, real-time benchmarking capabilities and other performance improvement methodologies toward the goal of enhancing patient outcomes and saving lives.

Geographic area and number and percentage of accountable entities and patients included: Since its initiation in 2003, participating hospitals have entered more than two million patient records into the GWTG-Stroke database. As of February, 2015, there are 2173 hospitals representing every state in the U.S, actively participating in GWTG-Stroke. Between 63.5%-83.8% (varies depending on the region of the country) of all ischemic stroke discharges in the U.S. are covered by GWTG-Stroke, In 2012, the latest year for which we have data, the total number of stroke discharges covered by GWTG-Stroke was 427,648.

2) Name of program and sponsor: GWTG-Stroke Hospital Recognition Program; Sponsor: American Heart Association/American Stroke Association

Purpose: This program confers awards on hospitals for success in using our programs (GWTG-Stroke and Target: Stroke) to improve quality of care for stroke patients. Participating in GWTG-Stroke is the first level of recognition. It acknowledges program

participation and entry of baseline data into the Patient Management Tool. Achievement Awards recognize hospitals that demonstrate at least 85 percent compliance in each of 7 Get With The Guidelines-stroke Achievement Measures. The different levels reflect the amount of time for which the hospital demonstrates performance:

- Bronze recognizes performance of 90 consecutive days.
- Silver recognizes performance of 12 consecutive months
- Gold recognizes performance of 24 consecutive months or more.

Silver Plus and Gold Plus Quality Awards are advanced levels of recognition acknowledging hospitals for compliance with Quality Measures embedded within the Patient Management Tool. A searchable database of hospitals awarded recognition is publicly available on the AHA website and every year a full page ad is placed in US News and World Report that lists all the hospitals that received recognition.
Geographic area and number and percentage of accountable entities and patients included: In FY 2015, 827 hospitals all across the country received GWTG-Stroke recognition awards, including many that care for underserved populations.

3) Name of program and sponsor: Target: Stroke; Sponsor: American Heart Association/American Stroke Association

Purpose: Target: Stroke is a national quality improvement initiative focused on improving acute ischemic stroke care by reducing door-to-needle times for eligible patients being treated with tPA.
Participating hospitals commit to reaching the Target: Stroke performance goal of 50 percent or more of eligible patients treated with thrombolytics within 60 minutes of hospital arrival. Ten key strategies are employed to meet this goal, including EMS pre-notification of hospitals, activating the stroke arrival team with a single call, rapid acquisition and interpretation of brain imaging, use of specific protocols and tools, premixing tPA, a stroke-team-based approach and rapid performance data feedback. Each hospital receives a detailed toolkit, including the 11 key strategies, protocols, stroke screening tools, order sets, algorithms, time trackers, patient education materials and other tools. Target: Stroke recognition is based entirely on meeting the performance goal of 50% on the Time to Thrombolytic measure. Each year, the names of all hospitals who are recognized by Target: Stroke for achieving the performance goal are published in U.S. News and World Report. Hospitals that exceed the performance goal and achieve 75% or more on this measure are also recognized. The list of hospitals receiving Target: Stroke recognition is publicly available via heart.org (See: http://www.heart.org/idc/groups/heart-public/@wcm/@gwtg/documents/downloadable/ucm_476918.pdf), along with a searchable database of hospitals who met the performance goal.
Target: Stroke Phase II aims to continue to reduce door-to-needle times for eligible patients being treated with tPA by establishing more aggressive goals for participating hospitals.

Geographic area and number and percentage of accountable entities and patients included: Target Stroke is a national program. In its first year (2010), more than 1,200 U.S. hospitals were enrolled. In FY 2015, 710 of the participating hospitals were recognized for achieving the Target: Stroke performance goal.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not applicable.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not applicable.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

The mean performance score from the Get with the Guidelines registry increased from 2012 to 2015 from 53% to 70%. This indicates that the implementation of the measure led to a positive trend in the proportion of patients receiving tPA treatment with a door-to-needle time of 60 minutes or less.

4b.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
Not applicable.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.
With regards to unintended consequences, reporting door to needle times has not led to unintended consequences of reduced use of tPA, which could happen if hospitals were neglecting to treat after 60 minutes to avoid a door to needle time <=60 minutes failure on the measure. In one study, that analyzes Get With The Guidelines-Stroke data, evidence that tPA use has increased over time, not decreased. More specifically, the data showed that for IV-rtPA the proportion of eligible patients who received treatment doubled over the 7-year period from 14.2% (2003) to 28.9% (2009), with a concomitant decline in the proportion of eligible patients who were not treated.(1) There was relatively little change in the proportion of patients with documented contraindications or missing data. These data patterns are consistent with improved care (ie, more eligible subjects treated).

Although there may be concerns that attempting to achieve shorter door-to needle times may lead to rushed assessments, dosing errors, and greater likelihood of complications, (2) there was no evidence in one study of worse in hospital outcomes or increased bleeding complications from tPA for patients with door-to-needle times <=60 minutes compared with those with door-to-needle times >60 minutes.(3) In-hospital mortality was lower among those patients treated in a more timely fashion, even after extensive risk adjustment. Lower mortality has not previously been reported with more timely tPA therapy within the first 3 hours after stroke onset, so this finding should be interpreted cautiously and should be replicated in independent data sets. This finding is, however, consonant with meta-analysis data indicating that late thrombolytic therapy beyond 3 hours after onset increases mortality and earlier thrombolytic therapy within 3 hours does not.(4-5) Importantly, the rate of symptomatic intracerebral hemorrhage in tPA-treated patients with door-to-needle times within 60 minutes was 4.7%, lower than the 5.6% rate of reported in patients with door-to-needle times >60 minutes.

These rates may also compare favorably with those observed in the NINDS trial (6.4%) and other phase IV studies.(6) These findings suggest that more rapid reperfusion therapy can be achieved without compromising short-term clinical outcomes.

Citations:

1. Reeves MJ, Grau-Sepulveda MV, Fonarow GC, Olson D, Smith EE, Schwamm LH. Are improvements in quality in the Get With The Guidelines–Stroke Program related to better care or better documentation? Circ Cardiovasc Qual Outcomes. 2011;4: 503–511. Available at: http://circoutcomes.ahajournals.org/content/4/5/503.full

2. Schwamm LH, Pancioli A, Acker JE III, Goldstein LB, Zorowitz RD, Shephard TJ, Moyer P, Gorman M, Johnston SC, Duncan PW, Gorelick P, Frank J, Stranne SK, Smith R, Federspiel W, Horton KB, Magnis E, Adams RJ. Recommendations for the establishment of stroke systems of care: recommendations from the American Stroke Association's Task Force on the Development of Stroke Systems. Stroke. 2005;36:690 –703. Available at: stroke.ahajournals.org/content/36/3/690.full

3. Wardlaw JM, Murray V, Berge E, del Zoppo GJ. Thrombolysis for acute ischaemic stroke. Cochrane Database Syst Rev. 2009;4:CD000213. Available at: http://stroke.ahajournals.org/content/35/12/2914.full

4. Lees KR, Bluhmki E, von Kummer R, Brott TG, Toni D, Grotta JC, Albers GW, Kaste M, Marler JR, Hamilton SA, Tilley BC, Davis SM, Donnan GA, Hacke W; ECASS, ATLANTIS, NINDS and EPITHET rt-PA Study Group, Allen K, Mau J, Meier D, del Zoppo G, De Silva DA, Butcher KS, Parsons MW, Barber PA, Levi C, Bladin C, Byrnes G. Time to treatment with intravenous alteplase and outcome in stroke:

an updated pooled analysis of ECASS, ATLANTIS, NINDS, and EPITHET trials. Lancet. 2010;375:1695–1703. Available at: http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(10)60491-6/fulltext

5. National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. N Engl J Med. 1995;333:1581–1587. Available at: http://www.nejm.org/doi/full/10.1056/NEJM199512143332401.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

0437 : STK 04: Thrombolytic Therapy

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**

Are the measure specifications completely harmonized?

No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

Measure #1952 assesses of the patients who received tPA within 4.5 hours, the percentage of patients who received tPA within the optimal time window of = 60 minutes. This measure focuses on the timely administration of tPA rather than whether or not the treatment should be administered. Data demonstrates that shortening door-to-needle times improves outcomes for acute ischemic stroke. Conversely, Measure #0437 assesses whether or not therapy was administered in eligible patients. As a result, the specifications differ where needed based on different populations and different focal points of the measure.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or

methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** 1952_Specifications_and_Data_Collection_Form_Appendix.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** American Heart Association/American Stroke Association
Co.2 **Point of Contact:** Melanie, Shahriary, melanie.shahriary@heart.org, 301-569-6159-
Co.3 **Measure Developer if different from Measure Steward:** American Heart Association/American Stroke Association
Co.4 **Point of Contact:** Melanie, Shahriary, melanie.shahriary@heart.org, 301-569-6159-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The below volunteers are part of the Get With The Guidelines-Stroke measures workgroup and are responsible for developing and maintaining measures included in the Get With The Guidelines-Stroke module.

*Eric E. Smith, MD, MPH, FRCPC
Chair
Assistant Neurologist
Massachusetts General Hospital

*Lee H. Schwamm, MD, FAHA
Vice Chairman of the Neurology Dept
Massachusetts General Hospital

*Gregg C. Fonarow, M.D., FACC
Professor of Medicine
Director, Ahmanson-UCLA Cardiomyopathy Center
Co-Director, UCLA Preventative Cardiology Program

Jeff Saver, MD, FAHA, FAAN
Professor of Neurology
Geffen School of Medicine at UCLA

Mathew Reeves, PhD, DVM
Associate Professor
Department of Epidemiology
Michigan State University

David Tong MD FAHA
Medical Director, CPMC Comprehensive Stroke Care Center
Director, CPMC Center for Stroke Research (CCSR)

Scott Kasner, MD, MSCE, FAHA
Professor of Neurology
Director, Comprehensive Stroke Center
University of Pennsylvania
Medical Center

Measures were reviewed by the GWTG- Exec Committee which includes the below volunteers as well as those above denoted with an asterisk:

Paul Heidenreich, MD,MS
Associate Professor of Medicine

Stanford University
VA Palo Alto Medical Center

Robert Berg, MD, FAAP, FAHA
Professor & Division Chief
Critical Care Medicine
Children's Hospital Philadelphia

Eric D. Peterson, MD, MPH, FAHA, FACC
Professor of Medicine
Vice Chair for Quality
Duke University Medical Center
Associate Director and Director of CV Research
Duke Clinical Research Institute

Adrian Hernandez, MD
Duke University Medical Cntr

Deepak L. Bhatt, MD, MPH, FAHA, FACC, FSCAI
Chief of Cardiology, VA Boston Healthcare System
Director, Integrated Interventional Cardiovascular Program, Brigham and Women's Hospital & VA Boston Healthcare System
Associate Professor of Medicine, Harvard Medical School

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2003

Ad.3 **Month and Year of most recent revision:** 09, 2014

Ad.4 **What is your frequency for review/update of this measure?** Annual Review

Ad.5 **When is the next scheduled review/update for this measure?** 10, 2017

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2111
Measure Title: Antipsychotic Use in Persons with Dementia
Measure Steward: Pharmacy Quality Alliance
Brief Description of Measure: The percentage of individuals 65 years of age and older with dementia who are receiving an antipsychotic medication without evidence of a psychotic disorder or related condition.
Developer Rationale: There is increasing concern about the overutilization of antipsychotics in older adults. Evidence shows that antipsychotic medications increase the risk of death and cerebrovascular events in people with dementia. This performance measure may help improve medication use and outcomes for older persons with dementia by reducing their exposure to potentially inappropriate medications through education of clinicians and patients on proper drug selection and usage. "Avoiding the use of inappropriate drugs is an important, simple, and effective strategy in reducing medication-related problems and adverse drug events in older adults." (1) Improvement in performance on this measure (reduction of non-indicated antipsychotic use in patients with dementia) may lessen the amount of cerebrovascular events and reduce the risk of death in elderly patients with dementia.

1.The American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc. 2012 Feb 29

Numerator Statement: The number of patients in the denominator who had at least one prescription and > 30 days supply for any antipsychotic medication during the measurement period and do not have a diagnosis of schizophrenia, bipolar disorder, Huntington's disease or Tourette's Syndrome.
Denominator Statement: All patients 65 years of age and older continuously enrolled during the measurement period with a diagnosis of dementia and/or two or more prescription claims within the measurement year for a cholinesterase inhibitor or an NMDA receptor antagonist within the measurement year where the sum of days supply is >60.
Denominator Exclusions: N/A

Measure Type: Process
Data Source: Administrative claims
Level of Analysis: Health Plan, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Mar 06, 2013 **Most Recent Endorsement Date:** Mar 06, 2013

Maintenance of Endorsement -- Preliminary Analysis

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the

prior evaluation.

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary from prior review in 2012

- FDA systematic reviews of 17 controlled trials presented by the developer for the previous evaluation includes: 2005) that led to the addition of the block box warning on atypical & typical antipsychotic in patients with dementia and risk of death.
- AHRQ systematic evidence review of the efficacy and comparative effectiveness of off-label atypical antipsychotic use.
- AGS/Beers clinical practice guideline (2012) on potentially inappropriate medication use, of which use of antipsychotics in patients with dementia was included. The guideline recommendation was graded.
- The developer also mentioned several other guidelines (e.g., from the AAN) that recommend non-pharmacologic approaches as the first-line treatment for behavioral problems in persons with dementia.
- Developer noted several situations in which the use antipsychotic medications could be appropriate in those with dementia.
- The developer cautioned that some patients with behavioral disturbances or agitation may not be adequately identified through claims data and may therefore not be excluded from the measure.

Changes to evidence from last review

☒ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☐ **The developer provided updated evidence for this measure:**

Guidance from the Evidence Algorithm

Process measure based on SR (Box 3)>QQC provided (Box 4)> Quantity: high; Quality: high; Consistency high (Box 5a)

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and voting on Evidence?*

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

Using CMS (2013) Part D contract data the mean measure rate was calculated. The number of measured entities is 731 contracts. Over 35 million Medicare beneficiaries were enrolled in prescription drug plans in 2013. The mean measure rate for all contracts was 12.8%. The range of scores are 7.7% at 10% to 19.4% at 90th.

- In the nursing home setting (2010), a study showed 43% of patients with dementia but no psychosis received the medication. Results from the MDS 2.0 (2010), showed the national average prevalence of antipsychotic use, in the absence of psychotic or related conditions, to be 18.5%. Those at high risk (High risk is defined as those residents who exhibit both cognitive impairment and behavior problems on the most recent assessment) were 39.4%.7 and those at low risk (Low risk is defined as all other residents who did not exhibit both cognitive impairment and behavior problems on the most recent assessment) were at 15.6%.
- A review by the OIG of atypical antipsychotic Medicare drug claims for elderly residents showed 14 percent of residents with Medicare claims for atypical antipsychotic drugs; 83 percent of the these claims were associated with prescribing for off-label conditions and 88 percent of the claims were associated with patients who had a diagnosis of dementia (a condition for which there is a black-box warning)
- Developer states using 2011 from large Medicare Advantage plans data showed that plans found that 13.7-15.9% of patients with dementia were receiving an antipsychotic medication, without evidence of a psychotic disorder or related condition.

Disparities

- From CMS (2013) Part D contract data, use of antipsychotic in persons with dementia is higher for those of lower economic status. Those with low income or on a subsidy were at 15.8% and those who were not were 11.3%.
- CMS contract data also shows use of antipsychotics in persons with dementia was higher for those residing in a nursing home facility longer than 100 days. For those living in the community the mean rate was 10.8% and those in a nursing home had a rate of 23.9%.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **This process measure relates directly to the overuse of anti-psychotic medications in patients with dementia. The data indicate definite over prescription with potential of serious side effects.

**AGS/Beers (2015) guideline updated since 2012- moderate evidence, strong recommendation

No new studies

High Evidence

1b. Performance Gap

Comments: **The mere presence of the problem strongly indicates a performance gap. Given the inherent limitations and errors in by dementia and psychosis diagnoses, there is still strong evidence of off-label use that is inappropriate and potentially harmful.

**performance gap demonstrates quality problem

1c. High Priority (previously referred to as High Impact)

Comments: **NA

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- Data source: Health plan medical and pharmacy claims and health plan member enrollment information.
- Administrative claims can be from multiple care settings include ambulatory and nursing homes and pharmacy claims.
- Level of analysis: prescription drug health plan, nation (population)
- Numerator: The number of patients in the denominator who had at least one prescription and > 30 days supply for any antipsychotic medication during the measurement period and do not have a diagnosis of schizophrenia, bipolar disorder, Huntington's disease, or Tourette's Syndrome.
- Denominator: All patients 66 years of age and older as of the last day of the measurement year who were continuously enrolled during the measurement year with both pharmacy and medical benefits and had a diagnosis of dementia and/or two or more prescription claims for a cholinesterase inhibitor or an NMDA receptor antagonist within the measurement year where the sum of days supply is >60.
- The goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent. The ICD-9 and ICD-10 codes (full listing and conversion is included in S.2b.

In the prior Neurology Endorsement Committee meeting(2012), some committee members questioned why Parkinson's disease was not excluded from the measure given that antipsychotic medications often are used appropriately for dementia-related psychosis in the patients. However, the Committee could not agree as to whether Parkinson's should be included or excluded.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing

Reliability testing from the prior review:

- Pilot testing of the measure was conducted by two large Medicare Advantage-Prescription Drug (MA-PD) plans and an additional employer-sponsored health plan using data from 2011. Two different plans had **4,288** and **6,323** members and the employer sponsored plan had **31,578** members qualified for the denominator.

Updated testing

- Developer provided updated testing using Centers for Medicaid & Medicare (CMS) 2013 Part D contract data. This data included both Medicare Advantage-Prescription Drug (MA-PD) plans and Medicare Prescription Drug Plans (PDP). Overall, 3,625,024 members qualitied for the denominator, 1,054,990 in MA-PD contracts, and 2,625,090 in PDP contracts.

SUMMARY OF TESTING

Reliability testing level ☒ **Measure score** ☐ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

Prior testing:

- Developer cited studies that evaluated the reliability and validity of pharmacy data.
- Developer described an approach that would assist in improving the recording of the diagnosis of dementia by using either a diagnosis of dementia or a medication marker for dementia (2 or more prescriptions claims and

>60 days supply for a cholinesterase inhibitor or an NMDA receptor antagonist). The two MA-PD plans recorded how often each of the requirements was met.
- NQF doesn't consider this to be reliability testing of the data elements

Updated testing:
- The developer fit a mixed effect logistic regression with varying intercept to test the hypothesis that the variation in measure rates across Medicare Part D contracts is statistically significant. This is equivalent to testing whether reliability =0.

Results of reliability testing

Prior testing results

See the discussion on data element testing in the Validity section below

- In the prior Neurology Endorsement Committee meeting(2012), some committee members questioned why Parkinson's disease a were not excluded from the measure given that antipsychotic medication s often are used appropriately for dementia-related psychosis in the patients. However, the Committee could not agree as to whether Parkinson's should be included or excluded.

Updated testing results:
- Developer presents the results that indicate that the rate variations at the contract level are statistically significant, as follows.

	Coefficient	Standard Error	Z	p-value
Intercept	-1.989	0.013	-152.72	<0.001
	Estimate	Standard Error	95% Confidence Interval	
Random Effects	0.302	0.011	0.281	0.324

The p-value for the likelihood ratio test was <0.001.

- These results demonstrate that reliability is not zero. However, they do not quantify the actual reliability value.
- Accuracy of the condition diagnoses in pharmacy claims has not been demonstrated (although if included in the cited studies, a summary would suffice). Score-level testing demonstrates only that the reliability is not zero, but does not quantify the reliability.

Guidance from the Reliability Algorithm :Specifications are clear and precise (Box 1)>empirical testing with statistical tests >see note about the lack of an actual reliability value.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

Rationale: Accuracy of the condition diagnoses in pharmacy claims has not been demonstrated (although if included in the cited studies, a summary would suffice). Score-level testing demonstrates only that the reliability is not zero, but does not quantify the reliability.

2b. Validity
Maintenance measures – less emphasis if no new testing data provided

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

Prior review:

- Developer presented data on the evaluation of comorbid psychoses diagnosis codes, dementia diagnosis codes and drug markers, and the use of prescription claims data.
- The developers assert that studies indicate that pharmacy claims are more complete than medical records and are of high quality.
- NQF guidance does not require additional data element reliability testing if validity of the critical data elements is demonstrated. It is unclear, however, whether any of the articles cited speaks to the accuracy of dementia or other psychosis diagnosis.
- The developer notes that the measure allows for identification of dementia patients through either dix codes or through medication codes. It is unclear, however, if the psychosis conditions included in the measure numerator are accurately reflected in pharmacy claims.
- Testing of this measure showed that 0.32% and 0.5% of patients with a diagnosis or medication marker for dementia also had a diagnosis for schizophrenia or bipolar
- Developer states that analysis shows that use of both dementia diagnosis codes and drug markers for dementia produce a more accurate representation of patients with dementia in the denominator.

Initially the measure was not endorsed because the validity of the measure as specified was adversely impacted because of the difficulties in identifying dementia patients and the fact that antipsychotic use in dementia patients is sometimes warranted. Additional information from the developer was submitted. The Committee agreed that the additional analysis submitted by the developer provided evidence that use of dementia medications as a way to identify dementia patient is a valid proxy. Upon re-vote, the Committee agreed the additional information provided by the developers was adequate to address their initial concerns about the validity of the measure

Updates to validity testing:

Face Validity:

Developer described a 2 steps process:

- The Pharmacy Quality Alliance (PQA's) Quality Metrics Expert Panel (QMEP) which contains members who have appropriate backgrounds. This group reviewed the measure prior to testing to ensure scientifically soundness and usefulness. The QMEP voted unanimously to recommend that PQA members consider the measure for endorsement.
- PQA members include organizations such as large pharmacy chains, health plans, quality organizations and pharmaceutical companies. At a meeting of PQA members (approximately 75 individuals), this group voted to endorse the measure, Antipsychotic Use in Persons with Dementia.
- The vote was 67% in favor of endorsement. Developer also stated that 3 different organizations who confirmed that the measure had face validity. No details were included to confirm this.
- The developer states it has not had results "rejected in whole or part," which NQF staff infers as face

validity at the performance score level.

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity only**
- ☐ **Empirical validity testing of the measure score**

Guidance from the Validity Algorithm:

Measures specification consistence with evidence (Box 1)> conducted using measure as specified **(**Box 2)>face validity assessed by experts (Box 4)>results of face validity by a majority of PQA members and other organizations (Box 5)> recommend moderate

Questions for the Committee:

- *Are psychosis conditions accurately portrayed in pharmacy claims so that those with conditions for which antipsychotic use may be appropriate are identified?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer states that this is not applicable

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- Developer states that the results indicate that the rate of variation at the contract level are statistically significant, which allows for discrimination between high performing plans and low performing plans (from reliability section).
- Updated Data from Medicare Part D data rates (2013) shows variation in rates from 7.73%- 14.16% with median of 12.1% and a mean of 12.83% and a standard deviation of 5.79%.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- Developer states the data sources are the health plan medical and pharmacy claims and health plan member enrollment information.
- Developer states that analysis presented that use of both dementia diagnosis codes and drug markers for dementia produce a more accurate representation of patients with dementia in the denominator.

2b7. Missing Data

- This measure is based on data collected through claims data that captures both diagnoses as well as medication use. The developer states that CMS calculates the measure for Part D plans and that the data is readily available through prescription claims data and medical data.

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **The specifications for validity are clear and direct.

2a2. Reliability Testing

Comments: **The validity testing is much improved by inclusion of prescription data, but the questions about diagnostic accuracy remain a concern. The proposal treats all anti-psychotics as equal without consideration whether some might be more appropriate for some off-label uses than others. The use of face validity as the only measure is a bit of a concern - if the experts who are the one's who are a part of the problem are also the ones evaluating the validity . . . ?

**face validity testing only - assessed by experts

2b2. Validity Testing

Comments: **Missing data are not a threat - but the variability in dementia and psychosis diagnosis are both a threat to the validity and to the effective implementation.

**some concern re-medication markers for dementia for identification of the denominator population because cholinesterase inhibitors/NMDA receptor antagonist used off label for pts with MCI/CI, migraines, ADHD

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **The presumption is that reliability testing has limitations and assuming validity measures are strong it can be assumed that the reliability is quite adequate.

**statistical testing insufficient

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- ALL data elements are in defined fields in electronic claims (Health plan medical and pharmacy claims and health plan member enrollment information). Administrative claims can be from multiple care settings, ambulatory, SNF and pharmacy.
- Coded by someone other than person obtaining original information
- The developer states that as a 2016 update, CMS calculates the measure for Part D plans. The data (prescription claims data and medical data) is readily available.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **Feasibility has the same problems with diagnostic accuracy. The goal is hard to argue with - the implementation is full of challenges. For example, if the drugs are often used as "chemical restraints" what are the options, when do they work, what are their side effects - are the as many and as bad as those from the anti-psychotic drug prescription? The fact that there have been no

adoptions since initial approval and thus no improvement data is a concern, despite having no argument with the goal.
**no concerns

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure [from OPUS]
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details
- The developer states that planned use includes CMS' addition of the measure to the Medicare Part D patient safety reports, beginning with year of service 2016, and addition to the 2018 Part D display measure set (using 2016 data).

Improvement results
- The developer stated they are not aware of any programs adopting the measure since its initial endorsement, and therefore does not have data demonstrating improvement.

Unexpected findings (positive or negative) during implementation: Developer did not identify any unexpected findings during implementation.

Potential harms: The developer did not identify any unintended consequences related to this measure.

Feedback
- This measure was reviewed by MAP for the Medicare Shared Savings Program in 2015. MAP noted that this measure addresses a critical program objective of including high-value measures in the set such as appropriate use measures. It also addresses a PAC/LTC core concept of inappropriate medication use. The measure promotes alignment as it is included in the MAP dual eligible beneficiary Family of Measures. MAP members reiterated that this measure addresses a large problem for persons with dementia.

***Questions for the Committee**:*
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **The relation between the measure and the improvement is unclear. Without any adoptions since initial approval the value of this measure is unknown (and perhaps the absence of adoptions is an indication). The best data on inappropriate

prescriptions is highly valuable and useable - the use of this measure to move those numbers is less clear.
**high usability

Criterion 5: Related and Competing Measures

Related or competing measures

Developer did not identify any related measures.

Harmonization

No

Pre-meeting public and member comments

Comment by Amy Elaine Sanders, MD
Organization American Academy of Neurology
Comment #5574: Medication overuse in this setting continues to be rampant, despite numerous measures requiring the opposite. There is little to no evidence to indicate efficacy, the black box warning is routinely ignored in clinical practice, and use of these meds is tantamount to chemical restraint. Strongly urge renewed endorsement. harmonization available with at least one other measure (AAN dementia set).

NATIONAL QUALITY FORUM

Measure Testing Evidence Attachment

NQF #: 2111 **NQF Project:** Neurology Project

1. IMPACT, OPPORTUITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT

Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance http://www.qualityforum.org/Measuring_Performance/Improving_NQF_Process/Evidence_Task_Force.aspxon evidence.

Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria. **(evaluation** http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx**criteria)**

a.1 Demonstrated High Impact Aspect of Healthcare: Affects large numbers, High resource use, Patient/societal consequences of poor quality

1a.2 If "Other," please describe:

1a.3 Summary of Evidence of High Impact *(Provide epidemiologic or resource use data)***:**
The scope of this measure is related to multiple high impact aspects of healthcare including affecting large numbers of patients and producing both high resource use and consequences of poor quality for patients.

Related to affecting large numbers of patients, the denominator focuses on patients 65 years of age and older with dementia. Current estimates describe dementia prevalence as affecting one in eight people age 65 and older, which equals about 13 percent or 5.2 million people.(1,2,3) Even more striking is that nearly half of people age 85 years and older are estimated to have this condition. (1,2,3) As the proportion of the U.S. population over age 65 continues to increase (especially with aging if the baby boom generation), the number of Americans with Alzheimer's disease and other dementias will increase as well.(2)

In addition, this measure focuses on medication safety or more specifically, the reduction of inappropriate medication use. Data has shown that about 90 percent of people 65 years of age and older take at least one medication, which is significantly more than any other age group.(4) Patient safety is a key aspect of quality related to medication use in the elderly, given their propensity to adverse drug events due to comorbid conditions and polypharmacy issues. Despite evidence of poor outcomes in older adults, inappropriate medications are prescribed and used as treatment.(5) Studies have shown that almost 30% of adverse drug events in primary care and 40% of adverse events in long-term care are preventable with problems mostly occurring at the initial ordering stage. (6,7) Total healthcare expenditures related to the use of potentially inappropriate medications has been estimated at $7.2 billion.(8)

Related to specifically to anti-psychotic drugs, their use is common in the elderly. A report by CMS in 2009 indicated that of the top 10 drugs paid for by Medicare Part D in 2006, 3 were atypical antipsychotic drugs.(9) In 2005, Medicaid spent more on atypical antipsychotic medications than on any other class of drugs, about $5.4billion. (10) In addition, a 2010 study published in the Archives of Internal Medicineshowed that over 30% of nursing home residents received at least one antipsychotic medication in 2006, and for over 30% of these patients there was no clinical indication for the medication.11 Related to financial consequences, a review of Medicare atypical antipsychotic drug claims for elderly nursing home residents showed fifty-one percent of the antipsychotic drug claims were erroneous (including not being used for medically accepted indications) amounting to

$116 million. (12)

Finally, serious safety concerns related to anti-psychotic use in the elderly are increasing. In particular, the health consequences of prescribing antipsychotic drugs for elderly patients with dementia are quite large, with side effects related to both increased morbidity (cardiovascular events such as heart attack and stroke) and risk of death. In 2005, the FDA issued an advisory requiring manufactures of atypical antipsychotic drugs to include a black-box warning. (13) The intent was to warn prescribers and consumers that the use of these drugs is not indicated in patients with dementia given the increased risk of mortality. A follow-up 2007 Agency for Healthcare Research and Quality (AHRQ) report which assessed off-label use of atypical antipsychotic drugs also found that all atypical antipsychotic drugs increase risk of death for elderly persons with dementia. (14)

1a.4 Citations for Evidence of High Impact cited in 1a.3:

1.Hebert LE, Scherr PA, Bienias JL, Bennett DA, Evans DA. Alzheimer's disease in the U.S. population: Prevalence estimates using the 2000 Census. Archives of Neurology 2003;60:1119–22.

2.Alzheimer's Association, 2012 Alzheimer's Disease Facts and Figures, Alzheimer's & http://www.alz.org/downloads/facts_figures_2012.pdf, accessed August 7, 2012.

3B.L. Plassman et al. Prevalence of Dementia in the United States: The Aging, Demographics, and Memory Study Neuroepidemiology 2007;29:125-132

4.Committee on Quality Health Care in America. Institute of Medicine. 2002. To err is human: building a safer health system. Washington, D.C: National Academy Press.

5.The American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc. 2012 Feb 29

6. Gurwitz JH, Field TS, Harrold LR et al. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. JAMA 2003;289:1107–1116.

7.Gurwitz JH, Field TS, Judge J et al. The incidence of adverse drug events in two large academic long- term care facilities. Am J Med 2005;118:251– 258.

8. Fu AZ, Jiang JZ, Reeves JH et al. Potentially inappropriate medication use and healthcare expenditures in the US community-dwelling elderly. Med Care 2007;45:472–476.

9 CMS, Data Analysis Brief: Medicare Part D Utilization Trends for Atypical Antipsychotics: 2006–2008, June 2009. Accessed at http://www.cms.hhs.gov on August 8, 2012.

10. Lagnado L. Nursing Homes Struggle To Kick Drug Habit. 2007. December 2007] http://online.wsj.com/article/SB119811286789841083.html. Accessed August 8, 2012.

11. Chen Y, Briesacher B, Field T, Unexplained Variation across U.S. Nursing Homes in Antipsychotic Prescribing Rates. Arch Intern Med. 2010 January 11; 170(1): 89–95

12. DHHS. Office of Inspector General, Medicare Atypical Antipsychotic Drug Claims for Elderly Nursing Home Residents. 2011. http://oig.hhs.gov/oei/reports/oei-07-08-00150.pdf Accessed August 7, 2012

13. Public Health Advisory: Deaths With Antipsychotics in Elderly Patients With Behavioral Disturbances, April 2005. Accessed at http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafety InformationforHeathcareProfessionals/PublicHealthAdvisories/ucm053171.htm on August 8, 2012.

14. AHRQ, Efficacy and Comparative Effectiveness of Off-Label Use of Atypical Antipsychotics (07-EHCOO3-EF), January 2007.

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:
There is increasing concern about the overutilization of antipsychotics in older adults. Evidence shows that antipsychotic medications increase the risk of death and cerebrovascular events in people with dementia. This performance measure may help improve medication use and outcomes for older persons with dementia by reducing their exposure to potentially inappropriate medications through education of clinicians and patients on proper drug selection and usage. "Avoiding the use of inappropriate drugs is an important, simple, and effective strategy in reducing medication-related problems and adverse drug events in older adults." (1) Improvement in performance on this measure (reduction of non-indicated antipsychotic use in patients with dementia) may lessen the amount of cerebrovascular events and reduce the risk of death in elderly patients with dementia.

1.The American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc. 2012 Feb 29

1b.2 Summary of Data Demonstrating Performance Gap *(Variation or overall less than optimal performance across providers)***: [***For Maintenance –*** *Descriptive statistics for performance results for this measure - distribution of scores for measured entities by quartile/decile, mean, median, SD, min, max, etc.***]**
In general, problems related to medication use are widespread. They bare significant costs in terms of both dollars and poor outcomes but are often preventable. In particular, research has shown usage of drugs for indications other than what the FDA approved the drug for (off-label use) is not untypical. For example one study published in 2006 showed off-label use accounted for just over 20 percent of prescriptions written in 2001. (1)

In relation to atypical antipsychotic drugs, a 2009 Department of Veterans Affairs study showed about 60 percent of individuals received antipsychotic drugs for off-label conditions.(2) An AHRQ report which looked at the drugs' efficacy and comparative effectiveness, listed the most common off-label uses as treatment of agitation in dementia, depression, OCD, PTSD, personality disorders, Tourette's syndrome, and autism.(3)

In the nursing home setting, a 2010 study published in the Archives of Internal Medicine showed that over 30% of nursing home residents received at least one antipsychotic medication in 2006 and 43% of patients with dementia and no psychosis received the medication.(4)

A review by the Office of Inspector General of atypical antipsychotic Medicare drug claims for elderly residents showed 14 percent of residents with Medicare claims for atypical antipsychotic drugs, 83 percent of the claims were associated with prescribing for off-label conditions and 88 percent of the claims were associated with patients who had a diagnosis of dementia (a condition for which there is a black-box warning).(5) In addition, a separate review was conducted by the Office of the Inspector General to understand how well nursing homes comply with extra protections set forth for nursing facility residents

receiving antipsychotic drugs.(6) The study looked for evidence of compliance with Federal requirements for resident assessments, documentation of decision making, care plan development and implementation. Strikingly, almost all records studied did not meet one or more Federal requirements for resident assessments and/or care plans.(6)

Finally, third quarter 2010 results from the Minimum Data Set (MDS) 2.0, which includes measures to facilitate nursing home resident assessment and care screening, showed the national average prevalence of antipsychotic use, in the absence of psychotic or related conditions to be 18.5%.(7) In addition, the national average prevalence of antipsychotic use, in the absence of psychotic or related conditions for those considered high risk was 39.4%.7 High risk is defined as those residents who exhibit both cognitive impairment and behavior problems on the most recent assessment. The national average prevalence of antipsychotic use, in the absence of psychotic or related conditions for those considered low risk was 15.6%.(7) Low risk is defined as all other residents who are not high risk (i.e., did not exhibit both cognitive impairment and behavior problems on the most recent assessment.)

Pilot testing of the measure under NQF endorsement consideration, Antipsychotic Use in Persons with Dementia, by two large Medicare Advantage plans using 2011 data also showed room for improvement in performance. Data across the 2 plans found 13.7-15.9% of patients with dementia were receiving an antipsychotic medication, without evidence of a psychotic disorder or related condition.(8) An additional analysis was conducted for a retiree population within an employer-sponsored health plan which found a rate of 18.5%.(8)

1b.3 Citations for Data on Performance Gap: [*For Maintenance* – *Description of the data or sample for measure results reported in 1b.2 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included***]**

1. D.C. Radley, S.N. Finkelstein, and R.S. Stafford, "Off-Label Prescribing Among Office-Based Physicians," Archives of Internal Medicine, Vol. 166, 2006, pp. 1021–1026.

2.D.L. Leslie, S. Mohamed, and R.A. Rosenheck, "Off-Label Use of Antipsychotic Medications in the Department of Veterans Affairs Health Care System," Psychiatric Services, Vol. 60, No. 9, 2009, pp. 1175–1181.

3.AHRQ, Efficacy and Comparative Effectiveness of Off-Label Use of Atypical Antipsychotics (07-EHCOO3- EF), January 2007.

4. Chen Y, Briesacher B, Field T, Unexplained Variation across U.S. Nursing Homes in Antipsychotic Prescribing Rates. Arch Intern Med. 2010 January 11; 170(1): 89–95.

5. DHHS. Office of Inspector General, Medicare Atypical Antipsychotic Drug Claims for Elderly Nursing Home Residents. 2011. http://oig.hhs.gov/oei/reports/oei-07-08-00150.pdf Accessed August 7, 2012.

6. DHHS. Office of Inspector General, Nursing Facility Assessments and Care Plans for Residents Receiving Atypical Antipsychotic Drugs. 2012. http://oig.hhs.gov/oei/reports/oei-07-08-00151.pdf. Accessed August 7, 2012.

7. CMS. MDS Quality Measure/Indicator Report.Psychotropic Drug Use- July/September 2010. http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MDSPubQIandResRep/qmreport.html Accessed August 14, 2012.

8. Pharmacy Quality Alliance Field Test Results, using 2011 data. www.pqaalliance.org

1b.4 Summary of Data on Disparities by Population Group: [*For Maintenance –Descriptive statistics for performance results for this measure by population group*]

Data is available to show disparities in antipsychotic prescribing relative to nursing home residence. A 2010 study published in the Archives of Internal Medicine reported evidence of facility-level variation in the prescribing of antipsychotics.1 The study also found newly-admitted nursing home residents were more likely to receive an antipsychotic if they were in a facility with a higher antipsychotic prescribing rate. This seems to signal that risky prescribing of antipsychotics seems to be a practice norm in some nursing homes and may be due to a nursing home antipsychotic prescribing culture.(1)

1b.5 Citations for Data on Disparities Cited in 1b.4: [*For Maintenance – Description of the data or sample for measure results reported in 1b.4 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*]

1. Chen Y, Briesacher B, Field T, Unexplained Variation across U.S. Nursing Homes in Antipsychotic Prescribing Rates. Arch Intern Med. 2010 January 11; 170(1): 89–95

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
The measure's focus is a process; the prescription and receipt of antipsychotic medications in patients with dementia. The link is to a health outcome; patients with dementia who receive an antipsychotic medication have been shown to be at increased risk for cardiovascular events and death.

1c.2-3 Type of Evidence *(Check all that apply)*:
Clinical Practice Guideline, Selected individual studies (rather than entire body of evidence), Systematic review of body of evidence (other than within guideline development)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
The body of evidence generally focuses on the efficacy and comparative effectiveness of antipsychotic medications for off-label uses in adults, and more specifically the efficacy, effectiveness and adverse effects (e.g., increased risk of death) of antipsychotics in patients with dementia. The body of evidence includes individual studies, several systematic evidence reviews & meta-analyses and evidence based guidelines.

There was a 2005 systematic evidence review by the FDA based upon the results of 17 placebo-controlled trials which led to the addition of the black box warning on atypical antipsychotic medications (to warn of the associated increased mortality risks for people with dementia). Since issuing that notification, the FDA has reviewed additional information that indicates the risk is also associated with conventional antipsychotics.

Reference:
Food and Drug Administration, Antipsychotics: conventional and atypical. http://www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm110212.htm. Accessed August 18, 2012

There was an AHRQ systematic evidence review of the efficacy and comparative effectiveness of off-label atypical antipsychotic use.

References:
Maher AL, et al. Efficacy and Comparative Effectiveness of Off-Label Use of Atypical Antipsychotics.

JAMA. 2011;306:1359-69.

Maglione M, Ruelaz Maher A, Hu J, Wang Z, Shanman R, Shekelle PG, Roth B, Hilton L, Suttorp MJ, Ewing BA, Motala A, Perry T. Off-Label Use of Atypical Antipsychotics: An Update. Comparative Effectiveness Review No. 43. (Prepared by the Southern California Evidence-based Practice Center under Contract No. HHSA290-2007-10062-1.) Rockville, MD: Agency for Healthcare Research and Quality. September 2011. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

There was also a 2012 systematic evidence review related to the AGS/Beers clinical practice guideline on potentially inappropriate medication use, of which antipsychotics in patients with dementia was included as one.

Reference:

The American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc. 2012 Feb 29

Article available online:
http://www.americangeriatrics.org/files/documents/beers/2012BeersCriteria_JAGS.pdf
Criteria and Evidence tables also available online:
http://www.americangeriatrics.org/files/documents/beers/2012AGSBeersCriteriaCitations.pdf

Meta-Analyses:

Schneider LS, Dagerman KS and Insel P. Risk of death with atypical antipsychotic drug treatment for dementia: meta-analysis of randomized placebo-controlled trials. JAMA 2005;294:1934-1943

Schneider LS, Dagerman K, Insel PS. Efficacy and adverse effects of atypical antipsychotics for dementia: meta-analysis of randomized, placebo-controlled trials. Am J Geriatr Psychiatry. 2006;14:191-210.

In addition, guidelines put forth from the American Academy of Neurology, the Third Canadian Consensus Conference on the Diagnosis and Treatment of Dementia & the California Workgroup on Guidelines for Alzheimer's disease Management all indicate that non-pharmacological strategies are the preferred first-line treatment approach for behavioral problems in patients with dementia.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles*)**:** About 25-30 studies.

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events*)**:** The evidence base used to support this measure is directly related to the scope of the measure; that is, studies have shown poor outcomes associated with antipsychotic use in patients with dementia. The strength of the body of evidence related to the inappropriate use of antipsychotics in patients with dementia has been determined to be high in terms of quality and strong in terms of strength. Based upon the

weight of the evidence, The FDA determined that there are increased risks of death in elderly patients with dementia-related psychosis observed with use of both conventional antipsychotics and atypical antipsychotics. Given this, the prescribing information for all antipsychotic drugs includes information about this risk in a Boxed Warning section of the medication. In addition, given the strength of the evidence, the American Geriatric Society has included antipsychotic medications for patients with dementia in its Beers list of Potentially Inappropriate Medications.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* There has been general consistency across studies in terms of magnitude and direction of effect.

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)***:**

Studies have shown that the risk of harms is higher than the potential benefit in terms of receipt of anti-psychotic medication for people with dementia. As an example of this, the FDA's analysis of 17 placebo-controlled trials that enrolled 5377 elderly patients with dementia-related behavioral disorders showed a risk of death of between 1.6 to 1.7 times that seen in placebo-treated patients.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? Yes

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: As indicated below the evidence was graded by AGS/Beers using the GRADE system and an interdisciplinary expert panel. The FDA and other guideline developers also graded the evidence associated with their reviews.

1c.11 System Used for Grading the Body of Evidence: GRADE

1c.12 If other, identify and describe the grading scale with definitions:

1c.13 Grade Assigned to the Body of Evidence: For AGS/Beers Drug Disease Interactions: Quality of Evidence: High, Strength of Evidence: Strong

1c.14 Summary of Controversy/Contradictory Evidence: There are differing opinions regarding which comorbid conditions may warrant the use of antipsychotics in patients with dementia. The conditions which were selected as appropriate, and therefore included in the numerator of this measure were done so based on a review of the evidence, harmonization of this measure's criteria with that of a similar (non NQF endorsed) CMS measure, and the advice of PQA's expert panel.

In addition, there may be some instances in which behavioral disturbances or agitation in patients with dementia may be treated unsuccessfully with non-pharmacologic method; therefore, pharmacologic treatment including anti-psychotics may be chosen as the next line of treatment. The measure relies on

medical and pharmacy claims data. Behavioral disturbances and agitation are not accurately captured with this type of data, and as such patients exhibiting these conditions could not be excluded from the measure. It should be noted though that the potential for improvement on the measure is large compared to the instances for which these exclusions might be necessary. In addition, this effect of this factor should not disproportionally impact one plan over another; rather the effect would likely be similar across health plans evaluated.

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:

Alexander GC, et al. Increasing off-label use of antipsychotic medications in the United States, 1995-2008. Pharmacoepidemiol Drug Saf. 2011;20(2):177-184

Briesacher BA, et al. The Quality of Antipsychotic Drug Prescribing in Nursing Homes. Arch Intern Med. 2005;165:1280-1285.

California Workgroup on Guidelines for Alzheimer's Disease Management. Guidelines for Alzheimer's disease management. Los Angeles, CA: Alzheimer's Disease and Related Disorders Association, Inc., Los Angeles Chapter. 2008.

Doody RS, Stevens JC, Beck C, et al. Practice parameter: management of dementia (an evidence based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2001;56:1154–66.

Dore DD, Trivedi AN, Mor V, Friedman JH and Lapane KL. Atypical antipsychotic use and risk of fracture in persons with Parkinsonism. Mov Disord 2009;24:1941-1948.

Food and Drug Administration, Antipsychotics: conventional and atypical. http://www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm110212. htm. Accessed August 18, 2012

Gill SS, et al. Antipsychotic drug use and mortality in older adults with dementia. Ann Intern Med. 2007;146(11):775-786.
Herrmann N, Gauthier S. Diagnosis and treatment of dementia: 6. Management of severe Alzheimer disease. CMAJ. December 2, 2008; 179(12): 1279 - 1287.

Huybrechts KF, et al. Differential risk of death in older residents in nursing homes prescribed specific antipsychotic drugs: population based cohort study. BMJ. 2012;344:e977.

Jeste DV, Lacro JP, Nguyen HA, et al. Incidence of tardive dyskinesia with risperidone versus haloperidol. Journal of American Geriatric Society. 1999;47:716–719.

Jeste D, Meeks T. To prescribe or not to prescribe? Atypical antipsychotic drugs in patients with dementia. South Med J. 2007;100:961–963.

Kales HC, et al. Mortality risk in patients with dementia treated with antipsychotics versus other psychiatric medications. Am J Psychiatry. 2007;164(10):1568-1576.

Liperoti R, et al. All-cause mortality associated with atypical and conventional antipsychotics among nursing home residents with dementia: a retrospective cohort study. J Clin Psychiatry. 2009;70(10):1340-1347.

Maher AL, et al. Efficacy and Comparative Effectiveness of Off-Label Use of Atypical Antipsychotics. JAMA. 2011; 306:1359-69.

Meeks & Jeste. Beyond the Black Box: What is The Role for Antipsychotics in Dementia? Curr Psychiatr. 2008 June 1; 7(6): 50–65.

Office of the Inspector General. Medicare atypical antipsychotic drug claims for elderly nursing home residents. 2011 (available at: http://oig.hhs.gov/oei/reports/oei-07-08-00150.asp).

Pollock BG, et al. A double-blind comparison of citalopram and risperidone for the treatment of behavioral and psychotic symptoms associated with dementia. Am J Geriatr Psychiatry. 2007;15:942–952

Schneeweiss S, et al. Risk of death associated with the use of conventional versus atypical antipsychotic drugs among elderly patients. CMAJ. 2007;176(5):627-632

Schneider LS, Dagerman KS and Insel P. Risk of death with atypical antipsychotic drug treatment for dementia: meta-analysis of randomized placebo-controlled trials. JAMA 2005;294:1934-1943

Schneider LS, Dagerman K, Insel PS. Efficacy and adverse effects of atypical antipsychotics for dementia: meta-analysis of randomized, placebo-controlled trials. Am J Geriatr Psychiatry. 2006;14:191-210.

Schneider LS, Tariot PN, Dagerman KS, et al. Effectiveness of atypical antipsychotic drugs in patients with Alzheimer´s disease. N Engl J Med 2006b;355:1525-1538.

Trifirò G, et al. All-cause mortality associated with atypical and typical antipsychotics in demented outpatients. Pharmacoepidemiol Drug Saf. 2007;16:538–544.

Vigen CL, et al. Cognitive effects of atypical antipsychotic medications in patients with Alzheimer's disease: outcomes from CATIE-AD. Am J Psychiatry. 2011;168(8):831-839.

Wang PS, et al. Risk of death in elderly users of conventional vs. atypical antipsychotic medications. N Engl J Med. 2005;353:2335–2341.

Wright RM, Roumani YF, Boudreau R, et al. Effect of central nervous system medication use on decline in cognition in community-dwelling older adults: findings from the Health, Aging And Body Composition Study. J Am Geriatr Soc 2009;57:243-250.

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)*:

Table 4, pg 5:
2012 American Geriatrics Society Beers Criteria for Potentially Inappropriate Medication Use in Older Adults

Organ System or Therapeutic Category or Drug: Antipsychotics, first (conventional) and second (atypical) generation

Rationale: Increased risk of cerebrovascular accident (stroke) and mortality in persons with dementia
Recommendation: Avoid use for behavioral problems of dementia unless nonpharmacological options have failed and patient is threat to self or others
Quality of Evidence: Moderate

Strength of Recommendation: Strong.

Table 3, pg 9:

2012 American Geriatrics Society Beers Criteria for Potentially Inappropriate Medication Use in Older Adults

Due to Drug–Disease or Drug–Syndrome Interactions That May Exacerbate the Disease or Syndrome

Disease: Dementia and cognitive impairment

Drug: Anticholinergics, Benzodiazepines, H2-receptor antagonists, Zolpidem, Antipsychotics, chronic and as-needed use

Rationale: Avoid because of adverse CNS effects. Avoid antipsychotics for behavioral problems of dementia unless nonpharmacological options have failed, and patient is a threat to themselves or others.

Antipsychotics are associated with an increased risk of cerebrovascular accident (stroke) and mortality in persons with dementia

Recommendation: Avoid

Quality of Evidence: High

Strength of Evidence: Strong

1c.17 Clinical Practice Guideline Citation: The American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc. 2012 Feb 29

1c.18 National Guideline Clearinghouse or other URL:
http://www.americangeriatrics.org/files/documents/beers/2012BeersCriteria_JAGS.pdf

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? Yes

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias: The American Geriatric Society appointed an 11-member interdisciplinary expert panel; any conflicts were disclosed.

1c.21 System Used for Grading the Strength of Guideline Recommendation: GRADE

1c.22 If other, identify and describe the grading scale with definitions:

1c.23 Grade Assigned to the Recommendation: Quality of Evidence Drug Disease Interactions: High, Strength of Evidence: Strong

1c.24 Rationale for Using this Guideline Over Others: This guideline was developed in 2012 specifically to update the previous Beers Criteria (which identifies potentially inappropriate medications) using a comprehensive, systematic review and grading of the evidence on drug-related problems and adverse drug events in older adults. The relationship of antipsychotic use to dementia is specifically addressed.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
2111_Evidence.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
There is increasing concern about the overutilization of antipsychotics in older adults. Evidence shows that antipsychotic medications increase the risk of death and cerebrovascular events in people with dementia. This performance measure may help improve medication use and outcomes for older persons with dementia by reducing their exposure to potentially inappropriate medications through education of clinicians and patients on proper drug selection and usage. "Avoiding the use of inappropriate drugs is an important, simple, and effective strategy in reducing medication-related problems and adverse drug events in older adults." (1) Improvement in performance on this measure (reduction of non-indicated antipsychotic use in patients with dementia) may lessen the amount of cerebrovascular events and reduce the risk of death in elderly patients with dementia.

1.The American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc. 2012 Feb 29

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
[2016 Entry]

The mean measure rate was calculated using all Centers for Medicaid & Medicare (CMS) 2013 Part D contract data. The number of measured entities is 731 contracts including 652 Medicare Advantage Prescription Drug Plan (MA-PD) contracts and 79 Medicare Prescription Drug Plans (PDP), Over 35 million Medicare beneficiaries were enrolled in prescription drug plans in 2013. This includes 22.5 million in Medicare Prescription Drug Plans and 12.8 million in Medicare Advantage Prescription Drug Plans.

Antipsychotic Use in Persons with Dementia Mean Measure Rate
All Contracts (N=731); Rate 12.8%
MA-PD (n= 652); Rate 12.8%
PDP (n=79); Rate13.1%

Minimum	0%
Maximum	48.0%
Mean	12.8%
Median	12.1%
Standard Deviation	5.8%
Interquartile Range	4.9%
Total Range	47.9%

Scores by deciles:

10th	20th	30th	40th	50th	60th	70th	80th	90th
7.7%	9.3%	10.3%	11.3%	12.1%	13.0%	14.0%	15.4%	19.4%

The rates reported in 1b.2 are the most recent scores. Since the measure has not been used prior to 2015, there is no data to compare rates over time. CMS intends to start reporting this measure to prescription drug plans in the 2018 Part D display measure

set (using 2016 data). The original testing information (2011 data) will be used to address improvement in section 4b.1.

[2012 Entry]

In general, problems related to medication use are widespread. They bare significant costs in terms of both dollars and poor outcomes but are often preventable. In particular, research has shown usage of drugs for indications other than what the FDA approved the drug for (off-label use) is not untypical. For example one study published in 2006 showed off-label use accounted for just over 20 percent of prescriptions written in 2001. (1)

In relation to atypical antipsychotic drugs, a 2009 Department of Veterans Affairs study showed about 60 percent of individuals received antipsychotic drugs for off-label conditions.(2) An AHRQ report which looked at the drugs' efficacy and comparative effectiveness, listed the most common off-label uses as treatment of agitation in dementia, depression, OCD, PTSD, personality disorders, Tourette's syndrome, and autism.(3)

In the nursing home setting, a 2010 study published in the Archives of Internal Medicine showed that over 30% of nursing home residents received at least one antipsychotic medication in 2006 and 43% of patients with dementia and no psychosis received the medication.(4)

A review by the Office of Inspector General of atypical antipsychotic Medicare drug claims for elderly residents showed 14 percent of residents with Medicare claims for atypical antipsychotic drugs, 83 percent of the claims were associated with prescribing for off-label conditions and 88 percent of the claims were associated with patients who had a diagnosis of dementia (a condition for which there is a black-box warning).(5) In addition, a separate review was conducted by the Office of the Inspector General to understand how well nursing homes comply with extra protections set forth for nursing facility residents receiving antipsychotic drugs.(6) The study looked for evidence of compliance with Federal requirements for resident assessments, documentation of decision making, care plan development and implementation. Strikingly, almost all records studied did not meet one or more Federal requirements for resident assessments and/or care plans.(6)

Finally, third quarter 2010 results from the Minimum Data Set (MDS) 2.0, which includes measures to facilitate nursing home resident assessment and care screening, showed the national average prevalence of antipsychotic use, in the absence of psychotic or related conditions to be 18.5%.(7) In addition, the national average prevalence of antipsychotic use, in the absence of psychotic or related conditions for those considered high risk was 39.4%.7 High risk is defined as those residents who exhibit both cognitive impairment and behavior problems on the most recent assessment. The national average prevalence of antipsychotic use, in the absence of psychotic or related conditions for those considered low risk was 15.6%.(7) Low risk is defined as all other residents who are not high risk (i.e., did not exhibit both cognitive impairment and behavior problems on the most recent assessment.)

Pilot testing of the measure under NQF endorsement consideration, Antipsychotic Use in Persons with Dementia, by two large Medicare Advantage plans using 2011 data also showed room for improvement in performance. Data across the 2 plans found 13.7-15.9% of patients with dementia were receiving an antipsychotic medication, without evidence of a psychotic disorder or related condition.(8) An additional analysis was conducted for a retiree population within an employer-sponsored health plan which found a rate of 18.5%.(8)

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

1. D.C. Radley, S.N. Finkelstein, and R.S. Stafford, "Off-Label Prescribing Among Office-Based Physicians," Archives of Internal Medicine, Vol. 166, 2006, pp. 1021–1026.

2.D.L. Leslie, S. Mohamed, and R.A. Rosenheck, "Off-Label Use of Antipsychotic Medications in the Department of Veterans Affairs Health Care System," Psychiatric Services, Vol. 60, No. 9, 2009, pp. 1175–1181.

3.AHRQ, Efficacy and Comparative Effectiveness of Off-Label Use of Atypical Antipsychotics (07-EHCOO3-EF), January 2007.

4. Chen Y, Briesacher B, Field T, Unexplained Variation across U.S. Nursing Homes in Antipsychotic Prescribing Rates. Arch Intern Med. 2010 January 11; 170(1): 89–95.

5. DHHS. Office of Inspector General, Medicare Atypical Antipsychotic Drug Claims for Elderly Nursing Home Residents. 2011. http://oig.hhs.gov/oei/reports/oei-07-08-00150.pdf Accessed August 7, 2012.

6. DHHS. Office of Inspector General, Nursing Facility Assessments and Care Plans for Residents Receiving Atypical Antipsychotic Drugs. 2012. http://oig.hhs.gov/oei/reports/oei-07-08-00151.pdf. Accessed August 7, 2012.

7. CMS. MDS Quality Measure/Indicator Report.Psychotropic Drug Use- July/September 2010. http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MDSPubQIandResRep/qmreport.html Accessed August 14, 2012.

8. Pharmacy Quality Alliance Field Test Results, using 2011 data. www.pqaalliance.org

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
[2016 Entry]
Low Income Subsidy
The measure was calculated at the contract level, and was grouped by low-income subsidy (LIS) status, which is a proxy for socioeconomic status. Medicare LIS is a subsidy paid by the Federal government to the drug plan for Medicare beneficiaries who need extra help with their prescription drug costs due to limited income and resources.

The same data source was used for this calculation as in 1b.2.

Antipsychotic Use in Persons with Dementia (N= number of contracts)
Measure Rate Statistics*
- All Contracts (N=731)
  - Mean 12.8%; Std Dev. 5.8%; Min 0%; Max 47.9%
- Low Income Subsidy (n= 726)
  - Mean 15.8%; Std Dev. 7.4%; Min 0%; Max 71.4%
- Non-LIS (n=630)
  - Mean 11.3%; Std Dev. 12.3%; Min 0%; Max 100.0%

Scores by deciles:

	10th	20th	30th	40th	50th	60th	70th	80th	90th
LIS	7.8%	10.6%	12.6%	14.3%	15.6%	17.2%	18.7%	20.6%	23.8%
Non LIS	0.0%	6.9%	8.1%	8.9%	9.7%	10.4%	11.4%	12.6%	15.0%

*Rates include outliers, which are often due to contracts with very small denominators.

Long-term Nursing Home Stay vs. Community Residence
The measure was calculated at the contract level, and was grouped by whether the patient resided in a nursing home for longer than 100 days at any time during the measure year versus whether they resided in the community during the year. The measure results show an increased use of antipsychotics in persons with dementia who reside in a nursing home facility longer than 100 days.

The same data source was used for this calculation as in 1b.2.

Antipsychotic Use in Persons with Dementia (N= number of contracts)
Measure Rate Statistics*
- All Contracts (N=731)
  - Mean 12.8%; Std Dev. 5.8%; Min 0%; Max 47.9%
- Community Only (N=731)
  - Mean 10.8%; Std Dev. 6.2%; Min 0%; Max 49.0%
- Long-Term Nursing Home Stay (N=678)
  - Mean 23.9%; Std Dev. 13.7%; Min 0%; Max 100.0%

*Rates include outliers, which are often due to contracts with very small denominators

[2012 Entry]
Data is available to show disparities in antipsychotic prescribing relative to nursing home residence. A 2010 study published in the Archives of Internal Medicine reported evidence of facility-level variation in the prescribing of antipsychotics.1 The study also found newly-admitted nursing home residents were more likely to receive an antipsychotic if they were in a facility with a higher antipsychotic prescribing rate. This seems to signal that risky prescribing of antipsychotics seems to be a practice norm in some nursing homes and may be due to a nursing home antipsychotic prescribing culture.(1)

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**

1. Chen Y, Briesacher B, Field T, Unexplained Variation across U.S. Nursing Homes in Antipsychotic Prescribing Rates. Arch Intern Med. 2010 January 11; 170(1): 89–95

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, High resource use, Patient/societal consequences of poor quality

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

The scope of this measure is related to multiple high impact aspects of healthcare including affecting large numbers of patients and producing both high resource use and consequences of poor quality for patients.

Related to affecting large numbers of patients, the denominator focuses on patients 65 years of age and older with dementia. Current estimates describe dementia prevalence as affecting one in eight people age 65 and older, which equals about 13 percent or 5.2 million people.(1,2,3) Even more striking is that nearly half of people age 85 years and older are estimated to have this condition. (1,2,3) As the proportion of the U.S. population over age 65 continues to increase (especially with aging if the baby boom generation), the number of Americans with Alzheimer's disease and other dementias will increase as well.(2)

In addition, this measure focuses on medication safety or more specifically, the reduction of inappropriate medication use. Data has shown that about 90 percent of people 65 years of age and older take at least one medication, which is significantly more than any other age group.(4) Patient safety is a key aspect of quality related to medication use in the elderly, given their propensity to adverse drug events due to comorbid conditions and polypharmacy issues. Despite evidence of poor outcomes in older adults, inappropriate medications are prescribed and used as treatment.(5) Studies have shown that almost 30% of adverse drug events in primary care and 40% of adverse events in long-term care are preventable with problems mostly occurring at the initial ordering stage. (6,7) Total healthcare expenditures related to the use of potentially inappropriate medications has been estimated at $7.2 billion.(8)

Related to specifically to anti-psychotic drugs, their use is common in the elderly. A report by CMS in 2009 indicated that of the top 10 drugs paid for by Medicare Part D in 2006, 3 were atypical antipsychotic drugs.(9) In 2005, Medicaid spent more on atypical antipsychotic medications than on any other class of drugs, about $5.4billion. (10) In addition, a 2010 study published in the Archives of Internal Medicine showed that over 30% of nursing home residents received at least one antipsychotic medication in 2006, and for over 30% of these patients there was no clinical indication for the medication.11 Related to financial consequences, a review of Medicare atypical antipsychotic drug claims for elderly nursing home residents showed fifty-one percent of the antipsychotic drug claims were erroneous (including not being used for medically accepted indications) amounting to $116 million. (12)

Finally, serious safety concerns related to anti-psychotic use in the elderly are increasing. In particular, the health consequences of prescribing antipsychotic drugs for elderly patients with dementia are quite large, with side effects related to both increased morbidity (cardiovascular events such as heart attack and stroke) and risk of death. In 2005, the FDA issued an advisory requiring manufactures of atypical antipsychotic drugs to include a black-box warning. (13) The intent was to warn prescribers and consumers

that the use of these drugs is not indicated in patients with dementia given the increased risk of mortality. A follow-up 2007 Agency for Healthcare Research and Quality (AHRQ) report which assessed off-label use of atypical antipsychotic drugs also found that all atypical antipsychotic drugs increase risk of death for elderly persons with dementia. (14)

1c.4. Citations for data demonstrating high priority provided in 1a.3

1.Hebert LE, Scherr PA, Bienias JL, Bennett DA, Evans DA. Alzheimer's disease in the U.S. population: Prevalence estimates using the 2000 Census. Archives of Neurology 2003;60:1119–22.

2.Alzheimer's Association, 2012 Alzheimer's Disease Facts and Figures, Alzheimer's & http://www.alz.org/downloads/facts_figures_2012.pdf, accessed August 7, 2012.

3B.L. Plassman et al. Prevalence of Dementia in the United States: The Aging, Demographics, and Memory Study Neuroepidemiology 2007;29:125-132

4.Committee on Quality Health Care in America. Institute of Medicine. 2002. To err is human: building a safer health system. Washington, D.C: National Academy Press.

5.The American Geriatrics Society 2012 Beers Criteria Update Expert Panel. American Geriatrics Society Updated Beers Criteria for Potentially Inappropriate Medication Use in Older Adults. J Am Geriatr Soc. 2012 Feb 29

6. Gurwitz JH, Field TS, Harrold LR et al. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. JAMA 2003;289:1107–1116.

7.Gurwitz JH, Field TS, Judge J et al. The incidence of adverse drug events in two large academic long-term care facilities. Am J Med 2005;118:251– 258.

8. Fu AZ, Jiang JZ, Reeves JH et al. Potentially inappropriate medication use and healthcare expenditures in the US community-dwelling elderly. Med Care 2007;45:472–476.

9 CMS, Data Analysis Brief: Medicare Part D Utilization Trends for Atypical Antipsychotics: 2006–2008, June 2009. Accessed at http://www.cms.hhs.gov on August 8, 2012.

10. Lagnado L. Nursing Homes Struggle To Kick Drug Habit. 2007. December 2007] http://online.wsj.com/article/SB119811286789841083.html. Accessed August 8, 2012.

11. Chen Y, Briesacher B, Field T, Unexplained Variation across U.S. Nursing Homes in Antipsychotic Prescribing Rates. Arch Intern Med. 2010 January 11; 170(1): 89–95

12. DHHS. Office of Inspector General, Medicare Atypical Antipsychotic Drug Claims for Elderly Nursing Home Residents. 2011. http://oig.hhs.gov/oei/reports/oei-07-08-00150.pdf Accessed August 7, 2012

13. Public Health Advisory: Deaths With Antipsychotics in Elderly Patients With Behavioral Disturbances, April 2005. Accessed at http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafetyInformationforHeathcareProfessionals/PublicHealthAdvisories/ucm053171.htm on August 8, 2012.

14. AHRQ, Efficacy and Comparative Effectiveness of Off-Label Use of Atypical Antipsychotics (07-EHCOO3-EF), January 2007.

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input was obtained.)*

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Behavioral Health, Behavioral Health : Serious Mental Illness, Mental Health, Mental Health : Serious Mental Illness, Neurology, Neurology : Cognitive Impairment/Dementia

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Safety, Safety : Medication Safety

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://pqaalliance.org/measures/default.asp

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Full_Listing_and_Conversion_Tables_ICD_9_to_10.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Addition of ICD-10 codes:
A description of the process used to identify ICD-10 codes is included in data field 2b2.2 in the Measure Testing Submission Form. The goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent. The ICD-9 and ICD-10 codes (full listing and conversion is included in S.2b.

Numerator: ICD-10 codes were added to Table Dementia D: Codes for Specific Psychotic Disorders or Related Conditions (Disease Codes to Identify Accepted Indications for Antipsychotic Medications).

Denominator: ICD-10 codes were added to Dementia Table A:Codes to Identify Dementia

Clarifying Language added to the Denominator:
Denominator: Clarifying language added to note that the prescription claims for a cholinesterase inhibitor or an NMDA receptor antagonist must be within the measurement year.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The number of patients in the denominator who had at least one prescription and > 30 days supply for any antipsychotic medication during the measurement period and do not have a diagnosis of schizophrenia, bipolar disorder, Huntington's disease or Tourette's Syndrome.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
The measurement year.

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The number of patients in the denominator who had at least one prescription and > 30 days supply for any antipsychotic medication during the measurement period (See Table Dementia C) and do not have a diagnosis for schizophrenia, bipolar disorder, Huntington's disease or Tourette's Syndrome (See Table Dementia D)

Table Dementia C: Antipsychotic Medications
Aripiprazole
Asenapine
Chlorpromazine
Clozapine
Fluphenazine
Haloperidol
Iloperidone
Loxapine
Lurasidone
Olanzapine
Paliperidone
Perphenazine
Pimozide
Quetiapine
Risperidone
Thioridazine
Thiothixene
Trifluoperazine
Ziprasidone
Note: The active ingredients are limited to oral, sublingual, injectable and intramuscular formulations only. Includes combination products.

Table Dementia D: Disease Codes for Specific Disorders for Exclusion
ICD-9
Schizophrenia:
295.0x to 295.9x
Bipolar/Manic Disorder:
296.0x
296.1x
296.4x to 296.9x
Huntington's disease
333.4
Tourette's Syndrome
307.23

ICD-10
Schizophrenia/schizophreniform
F20.0 F20.1 F20.2 F20.3
F20.5 F20.81
F20.89 F20.9 F25.9
Mania
F30.10 F30.11 F30.12 F30.13 F30.2
F30.3 F30.4 F30.8 F30.9
Bipolar
F31.0 F31.10 F31.11 F31.12 F31.13 F31.2
F31.30 F31.31 F31.32 F31.4 F31.5 F31.60 F31.61

F31.62 F31.63 F31.64 F31.70 F31.71
F31.72 F31.73 F31.74 F31.75
F31.76 F31.77 F31.78 F31.81 F31.89 F31.9
Tourettes
F95.2
Huntington's Disease
G10
Psychotic disorder
F06.0 F06.2 F06.33
Other psychotic disorders
F21 F23 F24 F28 F29 F53
Schizoaffective
F25.0 F25.1 F25.8
MDD with psychotic features
F32.3 F33.3

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
All patients 65 years of age and older continuously enrolled during the measurement period with a diagnosis of dementia and/or two or more prescription claims within the measurement year for a cholinesterase inhibitor or an NMDA receptor antagonist within the measurement year where the sum of days supply is >60.

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
All patients 66 years of age and older as of the last day of the measurement year who were continuously enrolled (i.e., had not disenrolled or died) during the measurement year with both pharmacy and medical benefits and had a diagnosis of dementia (Table Dementia A) and/or two or more prescription claims for a cholinesterase inhibitor or an NMDA receptor antagonist (Dementia Table B) within the measurement year where the sum of days supply is >60.

For a beneficiary for whom enrollment is verified monthly, the member may not have more than a 1-month gap in coverage (i.e., a member whose coverage lapses for 2 months [60 consecutive days] is not considered continuously enrolled).

Table Dementia B: Cholinesterase Inhibitors and NMDA Receptor Antagonists
donepezil
rivastigmine
galantamine
memantine
Note: The active ingredients are limited to oral and transdermal formulations only.

Dementia Table A: Codes to Identify Dementia
ICD-9
290.0
290.1x
290.2x
290.3
290.4x
294.10
294.20
331.0
331.82

ICD-10
F01.51 F02.80 F03.90

F05
G30.9
G31.83
A81.00 A81.01 A81.09
F01.50 F02.81 F03.91
F10.27 F10.96 F10.97
F13.27 F13.97
F18.97
F19.17 F19.27 F19.97
G30.0 G30.1 G30.8
G31.01 G31.09 G31.1

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
N/A

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
N/A

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
N/A

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
Step One:
Calculate the denominator by identifying the number of all eligible patients with either:
1) A diagnosis of dementia (Table Dementia A) and/or
2) Individuals with two or more prescription claims (within the measurement year) for a cholinesterase inhibitor or an NMDA

receptor antagonist (Table Dementia B) where the sum of days supply is >60

Step Two:
Calculate the numerator by identifying the number of persons in the denominator who have greater than 30 days supply for any antipsychotic medication during the measurement period (Table Dementia C) and do not have a diagnosis for schizophrenia, bipolar disorder, Huntington's Disease or Tourette's Syndrome (Table Dementia D).

Step Three:
Divide the numerator (step two) by the denominator (step one) and multiply times 100 to calculate the rate as a percentage.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Health Plan Medical and Pharmacy Claims. Health Plan member enrollment information.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other, Pharmacy
If other: The level of analysis for this measure is the prescription drug health plan, but it contains claims data from multiple care settings, including ambulatory, skilled nursing facility, pharmacy, etc.

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2111_MeasureTesting_MSF5_0_Data_updated011516.docx

NATIONAL QUALITY FORUM

Measure missing data in MSF 6.5 from MSF 5.0

NQF #: 2111	NQF Project: Neurology Project

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES
Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (evaluation criteria) Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See guidance on measure testing.
2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability*.)
2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*: Pilot testing of the measure was conducted by two large Medicare Advantage-Prescription Drug (MA-PD) plans using data for services provided in 2011. One plan had 105, 870 members with 81,639 members 65 years and older, and 4,288 members qualifying for the denominator (i.e. 65 years and older continuously enrolled, having either a diagnosis of dementia or a drug marker for dementia). The other MA-PD plan had 1,235,989 members, with 6323 members qualifying for the denominator. An additional analysis was conducted for a retiree population within an employer-sponsored health plan also using data from 2011. This plan had 440,088 members, all of which were 65 years and older. 31,578 members qualified for the denominator (i.e. they were 65 years and older continuously enrolled, having either a diagnosis of dementia or a drug marker for dementia). New Information January 2016 Additional testing was conducted using Centers for Medicaid & Medicare (CMS) 2013 Part D contract data. This data included both Medicare Advantage-Prescription Drug (MA-PD) plans and Medicare Prescription Drug Plans (PDP). There were 731 contracts (measured entities) in the analysis, which included 652 MA-PD contracts and 79 PDP contracts. Over 35 million Medicare beneficiaries were enrolled in prescription drug plans in 2013. This includes 22.5 million in MA-PD plans and 12.8 million in PDP plans. Overall, 3,625,024 members qualitied for the denominator, 1,054,990 in MA-PD contracts, and 2,625,090 in PDP contracts.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:

A.Testing was done to determine the reliability & validity of comorbid psychoses diagnosis codes (in the numerator) over the span of two years (the measurement year and year prior). This was done to understand whether there is under reporting of psychosis diagnoses over time (i.e., once someone has been diagnosed with psychosis, will it stop showing up in claims over time?)

B.Besides criteria for age and continuous enrollment, the denominator requires patients to have dementia. Given that dementia is under diagnosed and not always consistently recorded, the measure allows for either a diagnosis of dementia or a medication marker for dementia (2 or more prescriptions claims and >60 days supply for a cholinesterase inhibitor or an NMDA receptor antagonist). The two MA-PD plans recorded how often each of the requirements was met.

C.In addition, this measure is built on pharmacy claims data. Several studies have evaluated the reliability & validity of prescription claims data, including:

Kirking DM, Ammann MA, Harrington CA. Comparison of Medical Records and Prescription Claims Files in Documenting Prescription Medication Therapy," J Pharmacoepi. 1996, 5(1):3-15.

Choo PW, Rand CS, Inue TS, et al. Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. Med Care 1999;37:846-57.

Kwon A, Bungay KM, Pei Y, et al. Antidepressant use: concordance between self-report and claims records. Med Care 2003;41:368-74.

Saunders K, Simon G, Bush T, Grothaus L. Validation of pharmacy records in drug exposure assessment. J Clin Epidemiol 1997;50:619-25.

New Information January 2016

A mixed effect logistic regression was used with varying intercept to determine whether variation in measure rates across Medicare Part D contracts is statistically significant. Beneficiaries' numerator status (i.e., whether the beneficiary was in the numerator of the measure rate) was modeled based on the varying contract mean. A likelihood-ratio (LR) test was also performed to determine if a model with random effects would fit the data better than a standard logistic regression model without random effects.

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:

A.Results of the evaluation of comorbid psychoses diagnosis codes showed that of those patients in the numerator (taking antipsychotic medications without psychoses diagnosis in 2011), only 2.70% had diagnosis of schizophrenia or bipolar disorder prior to 2011.

The presence of schizophrenia in persons older than 65 years is very low. It is estimated that roughly 0.5% of people older than age 65 years have schizophrenia (Howard R, et al. Late-onset schizophrenia and very-late-onset schizophrenia-like psychosis: an international consensus. Am J Psychiatry 2000;157:172-8.) The testing of this measure showed that 0.32% and 0.5% of patients with a diagnosis or medication marker for dementia also had a diagnosis for schizophrenia or bipolar disease.

B.Results of the dementia code and medication marker analysis are as follows:

Number of subjects w/ diagnosis code for dementia:

MA-PD Plan 1: 2,574

MA-PD Plan 2: 1,898

Number of subjects w/ medication marker for dementia (2 or more rx, >60 ds)

MA-PD Plan 1: 3,121

MA-PD Plan 2: 5,758

Number of subjects with both diagnosis code for dementia and medication marker

MA-PD Plan 1: 1,411

MA-PD Plan 2: 1,311

Number of subjects that have either a diagnosis code or a drug marker for dementia (this is denominator)

MA-PD Plan 1: 4,288

MA-PD Plan 2: 6,323

This analysis shows that use of both dementia diagnosis codes and drug markers for dementia produce a more accurate representation of patients with dementia in the denominator. Further analysis by MA-PD plan 1 showed that only 53.72% of patients who took anti-dementia medications actually had diagnosis of dementia in claims data.

Data from two Medicare Advantage health plans showed the percentage of persons age 65 years older with continuous enrollment in the health plan and with a diagnosis code and/or a medication marker for dementia was 5.3% and 7.2%. These are lower rates than the estimate of 13% of persons age 65 and older having Alzheimer's disease and in line with evidence that dementia is underreported. The rates also seem reasonably valid given that Medicare Advantage plans tend to have mostly younger (<75 years) patients and fewer severely debilitated patients when compared to traditional Medicare (Parts A/B). Only 6% of Alzheimer's disease patients are aged 65 to 74 years (Alzheimer´s Association. 2012

citation 2 - 1a.4 Citations for Evidence)

C.The reliability and validity of prescription claims data has been evaluated in the literature. Kwon et. al. found high concordance between self-report and pharmacy claims data for anti-depressant medication use (agreement 85%, kappa .069). Most discordant cases could be resolved and not related to "errors" in self-report or claims data. Kirking et. al found in a study comparing prescription drug claims to medication use documented in medical records, that there were significantly more prescriptions documented in claims data as compared with corresponding medical records, which was even more apparent for high medication users vs. non-high users. A review by Lau et. al indicated that many studies have shown that pharmacy claims are more complete than medical records and are of high quality.

New Information January 2016

The results of the mixed effect logistic regression model are outlined in Table 1.

Table 1. Mixed Effect Logistic Regression Model, Antipsychotic Use in Persons with Dementia measure rate comparison across 731 Part D contracts

	Coefficient	Standard Error	Z	p-value
Intercept	-1.989	0.013	-152.72	<0.001
	Estimate	Standard Error	95% Confidence Interval	
Random Effects	0.302	0.011	0.281	0.324

The p-value for the likelihood ratio test was <0.001.

The standard deviation of the intercept term is different from 0 (0.302), as supported by the 95%CI, which we can interpret to mean that measure rates do vary at the contract level. Additionally, the likelihood ratio test shows that the varying intercept model (which allows measure rates to vary across contracts) fits the observed data better than a standard logistic regression model without random effects (which restricts all contracts to have the same average measure rate) with the p-value of <0.001, significant at alpha=0.05.

These results indicate that the rate variations at the contract level are statistically significant, which allows for discrimination between high performing plans and low performing plans. Based on these results, the measure is considered to be reliable.

2b. VALIDITY. Validity, Testing, including all Threats to Validity: H☐ M☐ L☐ I ☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* are consistent with the evidence cited in support of the measure focus (*criterion 1c)* and identify any differences from the evidence:

The measure's specifications are consistent with the evidence cited to support the measure (both in terms of the body of evidence & clinical practice guidelines). That is, the evidence shows that prescribing antipsychotic medications to elderly patients with dementia is potentially inappropriate given the increased risk of morbidity and mortality (i.e., poor outcomes). The measure looks at the percentage of dementia patients 65 years and older who are receiving an antipsychotic. In addition, given that some patients with dementia may have comorbid psychoses (for which the prescription of an antipsychotic may be warranted), the measure does account for this in the numerator.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

See section 2a2.1., Data Sample for Reliability Testing.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:

A. Testing was done to determine the reliability & validity of comorbid psychoses diagnosis codes (in the numerator) over the span of two years (the measurement year and year prior). This was done to understand whether there is under

reporting of psychosis diagnoses over time (i.e., once someone has been diagnosed with psychosis, will it stop showing up in claims over time?)

B. Besides criteria for age and continuous enrollment, the denominator requires patients to have dementia. Given that dementia is under diagnosed and not always consistently recorded, the measure allows for either a diagnosis of dementia or a medication marker for dementia (2 or more prescriptions claims and >60 days supply for a cholinesterase inhibitor or an NMDA receptor antagonist). The two MA-PD plans recorded how often each of the requirements was met.

C. In addition, this measure is built on pharmacy claims data. As mentioned in the reliability section, there have been several studies that examined the reliability & validity of prescription claims data.

Kirking DM, Ammann MA, Harrington CA. Comparison of Medical Records and Prescription Claims Files in Documenting Prescription Medication Therapy," J Pharmacoepi. 1996, 5(1):3-15.

Choo PW, Rand CS, Inue TS, et al. Validation of patient reports, automated pharmacy records, and pill counts with electronic monitoring of adherence to antihypertensive therapy. Med Care 1999;37:846-57.

Kwon A, Bungay KM, Pei Y, et al. Antidepressant use: concordance between self-report and claims records. Med Care 2003;41:368-74.

Saunders K, Simon G, Bush T, Grothaus L. Validation of pharmacy records in drug exposure assessment. J Clin Epidemiol 1997;50:619-25.

D. The measure has been tested for face validity (i.e., whether it appears to measure what it intends to measure) through review by PQA's Quality Metrics Expert Panel, PQA's full membership, and the health plans who pilot tested the measure.

New Information January 2016

Process to Identify ICD-9 to ICD-10 Conversion

Goal Statement: The goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.

Methods

To map ICD-9 to ICD-10, PQA staff used the following website: https://www.aapc.com/icd-10/codes/. This lookup tool is based on the CMS general equivalence mappings (GEMs): (http://www.cms.gov/Medicare/Coding/ICD10/2015-ICD-10-CM-and-GEMs.html). Although many codes in ICD-9-CM map directly to codes in ICD-10, in some cases, a clinical analysis was required to determine which code or codes should be selected. Key term searches for additional relevant codes not identified with the mapping process were conducted using the CMS ICD-10 lookup tool (http://www.cms.gov/medicare-

coverage-database/staticpages/icd-10-code-lookup.aspx). The CMS Chronic Condition Data Warehouse (CCW) for Alzheimer's Disease and related disorders and senile dementia (https://www.ccwdata.org/web/guest/home) was also consulted to determine completeness of the code list.

Measure Update Panel Review

The PQA Measure Update (MU) Panel reviewed the mapped codes and additional codes identified using key terms. Panel members also suggested additional codes based on the intent of the measure, for review. The MU Panel was instructed to review the codes to determine if the code maintained the intent of the measure. The MU Panel provided recommendations first via email, and then arrived at consensus following discussions over the course of 2 meetings.

Expert Opinion

Following the MU Panel recommendations, PQA staff sought additional expert advice to review the code list. Ryan Carnahan PharmD, MS, Associate Professor, University of Iowa College of Public Health, Department of Epidemiology, has significant expertise in this area. Dr. Carnahan served as a member on the National Quality Forum committee, Prioritizing Measure Gaps Project, Alzheimer Disease and Related Dementias. Dr. Carnahan reviewed the code lists. He was in agreement with the MU Panel's recommendation and provided a few additional suggestions.

Quality Metrics Expert Panel Review

The final step in the review process was to provide the recommendations of the MU Panel and Dr. Carnahan to the PQA Quality Metrics Expert Panel (QMEP). The QMEP's composition reflects PQA's membership and includes individuals with clinical or other technical expertise related to quality measurement. The QMEP is charged with:

- Reviewing recommendations from the MU Panel;
- Evaluating measure concepts proposed by PQA measure development teams;
- Reviewing comments from PQA members to determine necessary modifications to draft measures and/or variations to consider during testing;
- Reviewing the results of testing of draft measures; *and*
- Making final recommendations to the PQA membership regarding endorsement or retirement of measures.

The QMEP voted unanimously to approve the MU Panel/expert opinion recommendations for mapping ICD-9 to ICD-10 in the measure Antipsychotic Use in Persons with Dementia.

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:

A. Results of the evaluation of comorbid psychoses diagnosis codes showed that of those patients in the numerator (taking antipsychotic medications without psychoses diagnosis in 2011), only 2.70% had diagnosis of schizophrenia or bipolar disorder prior to 2011.

B. Results of the dementia code and medication marker analysis is as follows:

Number of subjects w/ diagnosis code for dementia:

MA-PD Plan 1: 2,574

MA-PD Plan 2: 1,898

Number of subjects w/ medication marker for dementia (2 or more rx, >60 ds)

MA-PD Plan 1: 3,121

MA-PD Plan 2: 5,758

Number of subjects with both diagnosis code for dementia and medication marker

MA-PD Plan 1: 1,411

MA-PD Plan 2: 1,311

Number of subjects that have either a diagnosis code or a drug marker for dementia (this is denominator)

MA-PD Plan 1: 4,288

MA-PD Plan 2: 6,323

This analysis shows that use of both dementia diagnosis codes and drug markers for dementia produce a more accurate representation of patients with dementia in the denominator. Further analysis by MA-PD plan 1 showed that only 53.72% of patients who took anti-dementia medications actually had diagnosis of dementia in claims data.

Information Updated January 2016 (information provided to NQF in November 2012)
The following italicized information regarding the validity of the measure was submitted to NQF in November of 2012, but was not included in the original testing document. We are adding this information here to ensure all information for the validity criteria is contained in the Measure Testing document:

Given that the medication markers for dementia are highly-specific to dementia, it is appropriate to use these markers to supplement the diagnosis codes for dementia for identification of the denominator population. When using the combination of medication marker and dementia diagnosis code, we found a fairly consistent rate dementia patients across the numerous Medicare contracts (average of 4.6% ; range of 3.4% to 5.9%). As noted earlier, the percentage of the population included in our dementia measure is not intended to replicate the overall rate of dementia in the general population since we are focused on a subset of dementia patients who do not have a diagnosis indicating psychoses or behavioral disturbance.

C. The reliability and validity of prescription claims data has been evaluated in the literature. Kwon et. al. found high concordance between self-report and pharmacy claims data for anti-depressant medication use (agreement 85%, kappa .069). Most discordant cases could be resolved and not related to "errors" in self-report or claims data. Kirking et. al found in a study comparing prescription drug claims to medication use documented in medical records, that there were significantly more prescriptions documented in claims data as compared with corresponding medical records, which was even more apparent for high medication users vs. non-high users. A review by Lau et. al indicated that many studies have shown that pharmacy claims are more complete than medical records and are of high quality.

D. Face Validity:

PQA's Quality Metrics Expert Panel (QMEP) (which contains members who have backgrounds in pharmacy, medicine, research, quality improvement and measures development) reviewed the measure prior to testing to ensure scientifically soundness and usefulness. The QMEP considered whether the age criteria, the inclusion of specific ICD-9 codes as appropriate indications for use of the antipsychotic, and stepwise construction of the measure calculation. The QMEP reviewed the results of the measure testing and found the measure to be feasible. Rates of antipsychotic use in persons with dementia varied between health plans testing the measure and there appeared to be significant room for improvement. The QMEP voted unanimously to recommend that PQA members consider the measure for endorsement.

PQA members (which include organizations such as large pharmacy chains, health plans, quality organizations and pharmaceutical companies, etc. See http://www.pqaalliance.org/members.htm for the full list) were notified prior to the PQA Annual Meeting in June 2012, of the opportunity to consider and vote for the performance measure during the meeting. Members received the measure description, key points and evidence, and measure specifications. During the PQA Business meeting, the measure was reviewed and there was open discussion of the measure. Nearly all of PQA members had a representative at the Annual meeting and were present for the vote (~75 of 88 member organizations present). The membership voted on the whether to endorse the measure, Antipsychotic Use in Persons with Dementia. The vote was 67% in favor of endorsement.

The measure was also reviewed and tested by 3 different organizations who confirmed that it had face validity. In addition, in particular, the plans were asked and confirmed in the positive after testing that they were able to accurately and easily identify the subset of patients who are in long term care facilities when they received the medications for dementia or psychoses.

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

N/A

2b3.2 Analytic Method (*Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:

N/A

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*

N/A

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

N/A

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:

N/A

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative*

assessment of relationship of risk factors to the outcome and differences in outcomes among the strata):

N/A

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment: N/A

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed and discriminated meaningful differences in quality.*)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

See information listed under section 2a2.1. Reliability Data Sample.

New Information January 2016

See information listed under section 2a2.1. Reliability Data Sample- New information December 2015

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:

The health plans were asked to pilot test the measure, generating the numerator, denominator and performance rate, along with further analysis to address additional research questions.

New Information January 2016

To assess significant differences in measure rates, we used 2013 Medicare Part D data for 736 plan contracts, and calculated the distribution mean, median, standard deviation, and interdecile range. These statistics are reported below in 2b5.3 Tables 1 and 2.

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:

1st MA-PD: Numerator: 588 Denominator: 4,288 Performance Rate: 13.70%

2nd MA-PD: Numerator: 1008 Denominator: 6,323 Performance Rate: 15.90%

Employer sponsored health plan: Numerator: 5,827 Denominator: 31,578 Performance Rate: 18.45%

Given the evidence (literature) indicates that use of antipsychotics in patients with dementia is linked to poor outcomes (increased morbidity and mortality), these pilot test results would indicate there is variation in care and room for improvement.

Information Updated January 2016 (information provided to NQF in November 2012)

The following italicized Information regarding the meaningful differences in performance of the measure was submitted to NQF in November of 2012, but was not included in the original testing document. We are adding this information here to ensure all information for the validity criteria is contained in the Measure Testing document:

For the 2011 Medicare data, the performance rate varied across the individual contracts from 10.2% to 20.3% with an average of 13.9% and standard deviation of 3.7%. Thus, there is variation in performance across the Medicare contracts

with some of the contracts having a rate that is nearly 2 standard deviations above the average.

New Information January 2016

Table 1. Variation in Measure Rates -2013 Medicare Part D data

Mean	Median	Standard Deviation
12.83%	12.1%	5.79%

Table 2. Interdecile Range of Measure Rates -2013 Medicare Part D data

Minimum	0.00%
10th Percentile	7.73%
20th Percentile	9.31%
30th Percentile	10.32%
40th Percentile	11.28%
50th Percentile	12.07%
60th Percentile	12.95%
70th Percentile	14.02%
80th Percentile	15.40%
90th Percentile	19.36%
Maximum	47.95%
Interdecile Range	14.16%

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:

See information listed under section 2a2.1. Reliability Data Sample.

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data*

sources specified in the measure):

Besides criteria for age and continuous enrollment, the denominator requires patients to have dementia. Given that dementia is under diagnosed and not always consistently recorded in the medical record, the measure allows for either a diagnosis of dementia or a medication marker for dementia (2 or more prescriptions claims and >60 days supply for a cholinesterase inhibitor or an NMDA receptor antagonist). The two MA-PD plans recorded how often each of the requirements was met.

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:

Number of subjects w/ diagnosis code for dementia:

MA-PD Plan 1: 2,574

MA-PD Plan 2: 1,898

Number of subjects w/ medication marker for dementia (2 or more rx, >60 ds)

MA-PD Plan 1: 3,121

MA-PD Plan 2: 5,758

Number of subjects with both diagnosis code for dementia and medication marker

MA-PD Plan 1: 1,411

MA-PD Plan 2: 1,311

Number of subjects that have either a diagnosis code or a drug marker for dementia (this is denominator)

MA-PD Plan 1: 4,288

MA-PD Plan 2: 6,323

This analysis shows that use of both dementia diagnosis codes and drug markers for dementia produce a more accurate representation of patients with dementia in the denominator. Further analysis by MA-PD plan 1 showed that only 53.72% of patients who took anti-dementia medications actually had diagnosis of dementia in claims data.

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: The measure is not currently stratified for disparities.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:

Although the measure is not currently stratified for disparities, disparities have been identified with regard to antipsychotic prescribing and nursing home status. In 2013 CMS will be requiring the use of an identifier to indicate whether patients are ambulatory or in a long term care facility. This identifier will aid in the stratification of the measure by these characteristics.

2.1-2.3 Supplemental Testing Methodology Information:
Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met? (*Reliability and Validity must be rated moderate or high*) Yes☐ No☐ Provide rationale based on specific subcriteria:
If the Committee votes No, STOP

New Information January 2016

SDS Risk Adjustment

- What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

  The Antipsychotic Use in Persons with Dementia measure is a process measure, and there is no conceptual basis for risk adjustment of sociodemographic variables. No SDS variables were analyzed for risk adjustment of this measure.

- Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk(*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

  N/A

- What were the statistical results of the analyses used to select risk factors?

  N/A

- Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

  N/A

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Other

If other: Prescription claims data.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Pilot test sites indicated the measure was feasible and results were able to be reported efficiently and accurately.

2016 update: CMS calculates the measure for Part D plans. The data is readily available (prescription claims data and medical data).

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

PQA develops and maintains numerous performance measures related to the medication use system. The Measures are the proprietary property of PQA, and it is in the interest of PQA to protect and promote the appropriate use of the Measures. PQA may approve an organization´s use of the Measures; however, no organization may use the Measures without first obtaining permission from PQA prior to using the Measures. Certain uses of the Measures are only approved with a licensing agreement from PQA that specifies the terms of use an the licensing fee. PQA reserves the right to determine the conditions under which it will approve and/or license the Measures.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Regulatory and Accreditation Programs Quality Improvement (Internal to the specific organization)	Quality Improvement with Benchmarking (external benchmarking to multiple organizations) CMS Medicare Part D- Patient Safety Reports http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/index.html

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Name of program and sponsor: CMS Medicare Part D Drug Benefit
Purpose: Decrease use of antipsychotics in Part D beneficiaries with dementia
Geographic area: National, approximately 38 million beneficiaries in Part D plans.

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

CMS has been considering using this measure in the Medicare Part D prescription drug program. CMS´ further evaluation of the measure has now been completed and the measure will be used N/in the display measure set using 2016 data.

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Planned use includes CMS' addition of the measure to the Medicare Part D patient safety reports, beginning with year of service 2016, and addition to the 2018 Part D display measure set (using 2016 data). From the Memo (Amy Larrick, Acting Director, Medicare Drug Benefit and C&D Data Group, Request for Comments: Enhancements to the Star Ratings for 2017 and Beyond:

CMS (Medicare Part D) will develop new patient safety APD (Antipsychotic Use in Persons with Dementia) measure reports to provide to Part D sponsors on a monthly basis through the Patient Safety Analysis website beginning with year of service 2016. CMS also recommends adding the overall APD measure plus breakout rates for community-only residents, short-term nursing home residents, and long-term nursing home stay residents to the 2018 Part D display measure set (using 2016 data) to continue to draw attention to the inappropriate use of antipsychotics in persons with dementia without an appropriate mental health diagnosis in both the community and nursing home settings. The APD measure will replace the Rate of Chronic Use of Atypical Antipsychotics by Elderly Beneficiaries in Nursing Homes display measure.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance

results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

N/A

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
PQA is not aware of any programs adopting the measure since its initial endorsement, and therefore we do not currently have data demonstrating improvement. However, per the CMS memo dated November 12, 2015, "Medicare Drug Benefit and C&D Data Group, Request for Comments: Enhancements to the Star Ratings for 2017 and Beyond", CMS will use this measure to monitor use of antipsychotics in persons with dementia covered by the Medicare drug benefit. CMS' most recent analyses show that there has been little change in the measure rate since initial endorsement of the measure in 2012. This helps support the need for a heightened focus and opportunity to decrease the use of antipsychotics in person with dementia and supports CMS´ decision to adopt the measure into the Part D Star Ratings program.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
2016 Update
No unintended negative consequences were identified during the additional testing of the measure.

2012 Entry
As referenced previously, this measure is built on medical and pharmacy claims data. There have been several studies that validate the reliability & validity of prescription claims data.
In addition, additional analyses were carried out during pilot testing (refer to results in sections 2a2.3, 2b2.3, 2b5.3, 2b6.3) to confirm the consistency and accuracy of the measure.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications completely harmonized?

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2863
Measure Title: CSTK-06: Nimodipine Treatment Administered
Measure Steward: The Joint Commission
Brief Description of Measure: Proportion of subarachnoid hemorrhage (SAH) patients age 18 years and older for whom nimodipine treatment was administered within 24 hours of arrival at this hospital.

This is the sixth measure in a set of measures developed for Joint Commission Comprehensive Stroke Certification. The other measures in the set include CSTK-01 National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients; CSTK-02 Modified Rankin Score (mRS) at 90 Days; CSTK-03 Severity Measurement Performed for Subarachnoid Hemorrhage (SAH) and Intracerebral Hemorrhage (ICH) Patients (Overall Rate). Although it is not required that these measures are reported in conjunction with each other, The Joint Commission develops measures in sets in order to provide as comprehensive a view of quality for a particular clinical topic as possible.
Developer Rationale: Cerebral vasopasm is a serious complication following subarachnoid hemorrhage (SAH), occurring in 30% to 70% of patients and accounting for nearly 50% of the deaths in patients surviving to treatment (Bederson, 2009). Constriction of the arterial lumen results in diminished cerebral perfusion distal to the affected artery, which produces a delayed neurological deficit that may progress to cerebral infarction without early management of the ruptured aneurysm. The arterial narrowing that occurs in cerebral vasospasm is typically a transient or temporary event, lasting from a few days up to 3 weeks.
The main goal of current treatment is to prevent or limit the severity of cerebral vasospasm. Only two treatments are generally accepted as proven and valuable for the prevention of ischemic stroke and reduction of ischemic complications:

- Treatment with cerebroselective calcium channel blocker nimodipine-Nimotop (60mg po q4h for 21 days after hemorrhage or after hospital discharge if discharged within 21 days) (Leifer, 2011);
- Aggressive hypervolemic, hypertensive, hemodilution therapy (i.e., triple-H therapy) with pressor agents and volume expansion (colloids) while monitoring the central venous pressure (CVP) or pulmonary capillary wedge pressure (PCWP), following early clipping of the aneurysm.

Numerator Statement: SAH patients for whom nimodipine treatment was administered within 24 hours of arrival at this hospital.
Denominator Statement: SAH patients
Denominator Exclusions:
- Patients less than 18 years of age
- Patients who have a Length of Stay greater than 120 days
- Patients with Comfort Measures Only documented on the day of or day after hospital arrival
- Patients enrolled in Clinical Trials
- Patients discharged within 24 hours of arrival at this hospital

Measure Type: Process
Data Source: Electronic Clinical Data, Paper Medical Records
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

The developer provides a diagram of the relationship of this process of care (administration of Nimodipine) to patient outcomes (fewer ischemic deficits and improved neurological outcome). The developers reference two sources for evidence of this relationship:

- **2012 Guidelines for the management of aneurysmal subarachnoid hemorrhage(aSAH)** from the American Heart Association/American Stroke Association for " Management of Cerebral Vasospasm and DCI After aSAH: Class I Recommendation (Procedure/treatment SHOULD be performed/administered) "*Oral nimodipine should be administered to all patients with aSAH*." Level of Evidence: A (Data derived from multiple randomized clinical trials or meta-analyses).
- **2007 Cochrane Systematic Review** "Calcium antagonists for aneurismal subarachnoid haemorrhage" reported that "Overall quality of evidence across 16 studies, involving 3361 patients, appears to be high" and concluded that "the outcome after subarachnoid hemorrhage, in terms of survival and being independent in activities of daily living, is improved by treatment with calcium channel blockers (antagonists). If the largest trial (Haley 1993, n= 906) is excluded from the analysis, the results are no longer statistically significant, and therefore the evidence is not beyond all doubt. However, given the high likelihood of benefits and the modest risks associated with this treatment, the review authors concluded that calcium antagonists, in the form of oral nimodipine 60 mg every four hours, are useful in patients with subarachnoid hemorrhage from a ruptured aneurysm." The aggregate results found a lower relative risk for poor outcome (RR=0.81 (CI 0.72 – 0.92)) and secondary ischemia (RR=0.64 (CI 0.49-0.83)).

Exception to evidence – not applicable

Guidance from the Evidence Algorithm

Process measure/systematic review (Box 3) → QQC presented (Box 4) → Quantity: high; Quality: high; Consistency high(Box 5) → Box 5a High

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

Preliminary rating for evidence: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provided the following data for current performance:

	Pilot test (2010-2013)	1Q 2015	2Q 2015
# sites	66	39	51
# patients (den)	1229	572	878
average rate	71%	86% (hospital mean rate)	83% (hospital mean rate)
range	0 -100%	10th percentile = 67% 90th percentile rate = 97%	10th percentile = 75% 90th percentile rate = 92%
National aggregate rate		86%	81%

Disparities

2005-2010 data from the Nationwide Inpatient Sample - multivariable analyses identified race/ethnicity as a significant predictor of both inpatient mortality (p=0.003) and discharge to institutional care (p < 0001). The study found that Hispanic patients were the least likely to have a poor outcome, and Asian/Pacific Islander patients were most likely to have a poor outcome.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Should this measure be used to assess disparities in care?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **This is a process measure. Agree with the preliminary rating of high. The quality of evidence is high that administration of nimodipine according to this protocol is associated with increased independence and decreased mortality. The quantity of evidence is high with 16 studies, although most of the evidence comes from the largest study.

**The process measure is clear and well supported. The outcome cost/benefit is less clear. There is an indicated $3.50 cost per chart. I have some doubts as to whether a formal measure focus will produce sufficient improvement in outcome data to justify the training, data collection and reporting costs.

1b. Performance Gap

Comments: **Performance data was provided, showing an 83-86% uptake rate on this recommendation. However, the 10th percentile uptake was still high (67-75%), indicating that while improvement is still possible, huge gains are unlikely. While there is a racial disparity in outcomes, it is not large. Agree with assessment of moderate for this criterion.

**I see data indicating a 14%-30% performance gap. Disparities were also present consistent with many other disparities.

1c. High Priority (previously referred to as High Impact)

Comments: No comments provided

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Medical record abstraction (electronic or paper). This is not an eMeasure.

Specifications:

- Six data elements define the numerator and seven data elements define the denominator.
- Denominator diagnosis is defined by ICD-10 codes. An attached spreadsheet identifies twenty ICD-10 codes for non-traumatic subarachnoid hemorrhage (presumably from a ruptured aneurysm).
- Sampling monthly or quarterly is allowed.
- The measures is specified at the hospital level of analysis.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

Inter-rater reliability testing of the data elements was conducted at twelve sites with 281 total records. Percent agreement and Kappa scores were used to compare two sets of abstractors. This is an appropriate test for data element reliability for abstracted data.

Results of reliability

The percent agreement for the six numerator data elements was greater than 95% except for the Admitting time at 82%. A Kappa score was calculated only for the Nimodipine administration equals 0.93 (Landis and Koch (1977) interpretation of Kappa: 0.81-1.00 =Almost perfect agreement)

Guidance from the Reliability Algorithm

Precise specifications (Box 1) → empirical testing with statistical tests as specified (Box 2) → reliability testing with patient-level data elements (Box 8) → assessed all critical data elements (Box 9) → high/moderate certainty the data elements are reliable (Box 10) → Moderate

Note: When only data element testing is performed (no measure score testing) the highest rating possible is moderate.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence:

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing:

- Empirical testing of the measure score:
  Hypothesis 1: Hospital results for two process measures for hemorrhagic stroke should be positively correlated. (CSTK 03: Severity Measurement Performed for Subarachnoid Hemorrhage (SAH) and Intracerebral Hemorrhage (ICH) Patients (Overall Rate) and CSTK 06:Nimodipine Treatment Administered. A
  Pearson Correlation Coefficient, an appropriate statistic for this purpose, was calculated to compare the results of these two measures.
  Hypothesis 2: Hospitals that do well on one stroke measure are likely to do well on other stroke measures. Pearson Correlation coefficients were calculated to compare results of several stroke measures.
- Face validity of data elements was reported but did not address face validity of the measure score as a representation of quality as required for the criterion.
- Data element validity was assessed for accuracy and clarity by hospitals but the data element validity criterion requires comparison to a gold standard.

Validity testing results:

Hypothesis 1: The developers interpret the results as " demonstrates that the CSTK06 rate was negatively and slightly correlated to the CSTK03 measure. Because the data is not normally distributed, the correlation coefficient on correlation may not be the best measure of association."

Pearson Correlation Coefficients, N = 66 Prob > |r| under H0: Rho=0

	CSTK03	CSTK06
Correlation Coefficient	-0.02229	1.00000
P value	0.8590	

The results do not confirm the hypothesis.

Hypothesis 2: The developers present a correlation table of several stroke measures and interpret the results as "shows a positive correlation and statistical significance which indicates that hospitals with high quality on one CSTK measure tend to have high correlations on the other stroke measures."

The results found that this measure has the highest, significant correlation with CSTK 01 National Institutes of Health

Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do you agree with the developer's interpretation of the validity test results?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:
The developers provide frequencies of exclusions for the pilot test and the 1Q and 2Q 2015 cohorts.
In 42 hospitals the frequency of all exclusions except for Comfort Measures Only (CMO) and Discharged within 24 hours are less than 1%. For this measure the frequency of CMO was 2% (range 0.5-6%) and for Discharged within 24 hours is 64% (range 48 – 100%).

Questions for the Committee:

- *In this sample more than half of the patients are excluded. Does the large number of exclusions for Discharged within 24 hours pose a threat to validity of the measure? Is this a reasonable exclusion for the measure?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful differences *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

Descriptive statistics for measure: N=66 hospitals
Overall rate=60.1% Standard Deviation=31.1%
Minimum=0% Median=70% Maximum=100%

The range in hospital performance indicates substantial variability to distinguish providers.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*
- *Is there enough variation that audiences can identify high quality providers from lower quality providers?*

2b6. Comparability of data sources/methods: not applicable

2b7. Missing Data
The developer describes the handling of missing data in the measure specifications (S.22) but does not provide information on the frequency of missing data or potential impact on measure results.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → potential threats to validity mostly assessed (Box2) →empirical validity testing (Box 3) → measure score testing (Box 6) → sound conceptualization and hypotheses (Box 7) →moderate confidence that score are a valid indicator of quality(Box 8b) → Moderate

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **As with many measures it is difficult to assess the value of a formal measure against the prevailing current best practice implementation.

I was particularly puzzled by a statement that Nimodipine was not demonstrated as effective in presenting vasospasm, but this is a major rationale for this measure.

2a2. Reliability Testing

Comments: **It would appear that the two measures for subarachnoid and intercerebral hemorrhage severity scoring and nimodipine administration in SAH should be positively correlated; therefore, that they have no relationship is a bit surprising.

The correlation table provided shows a correlation between the various stroke measures of quality. This seems to provide some indication of validity although the sample sizes for each of the measure is only ~42 individuals.

**Sample sizes for validity testing were also quite adequate - though, as indicated above I have some question about the strength of the validity.

2b2. Validity Testing

Comments: **The major threat to validity is that half of the sample size is excluded by the exclusion criterion of discharged within 24 hours. Although it seems necessary to exclude these individuals because the actual quality measure is starting nimodipine within 24 hours, this exclusion severely limits the ability of the measure. Patients who are discharged within the 24 hours should have a home prescription for nimodipine and most should be started while they are inpatients, even though they may be discharged. Would it be possible to assess the discharge prescriptions for nimodipine? With a range of 0-100% implementation of nimodipine prescribing, it would seems that there is enough difference to detect differences in quality, though with appropriate education of prescribers, an institution can rapidly change the compliance with this indicator as indicated by the difference in the pilot and implementation numbers. It would be helpful to know how often data was missing in the pilot and first two quarters of implementation. Agree with ranking of validity as moderate.

**No notable missing data.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Although a number of hospitals were included in the testing, only 281 records were used. This number seems a little low. The mix of hospitals does seem to provide a broad basis for reliability. The inter-rater reliability was high. Agree with ranking of moderate.

**The sample size and reliability testing seemed quite adequate.

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is medical record abstraction: not all hospitals are able to generate the data electronically so a paper-based option is available
- Some data elements are in electronic form
- The measure is in use in the Joint Commission's Stroke Certification program.
- Data collection burden is significant:
  - Medical record review to abstract all data elements required for the measure set averaged 45 minutes per record.
  - Developer estimated " that the cost per case to abstract for this measure was approximately $3.50."

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Is the burden of data collection reasonable for a national performance measure?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy

Comments: **While most of the data elements are available and routinely collected during clinical care. The two elements of comfort measures only and clinical trial participation would have to be hand-extracted many times. The pilot testing demonstrated a 45-minute time for the data extraction which seems significant and would, perhaps, lead to less implementation. The strategy, however, seems sound. Agree with ranking as moderate.
**The data should be relatively easily generated during care and eventually found in EHRs. At the same time I have some concern about the training time, data entry and data monitoring costs.

Criterion 4: Usability and Use
Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact /improvement and unintended consequences

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

The measure is currently used for internal quality improvement for Care Certification for Comprehensive Stroke Centers. Planned uses include public reporting and external benchmarking to other organization – no timeframe noted.

Accountability program details none

Improvement results The developers note and improvement from the pilot test (average results 70%) to the 1q and 2Q 2015 data with mean hospital results around 80-85%.

Unexpected findings (positive or negative) during implementation None reported

Potential harms None reported

Feedback : new measure with little use

***Questions for the Committee**:*

- o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Are hospitals likely to improve results quickly to overall high performance?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments: **While the results of this measure are not meant for public viewing, they can be indirectly assessed by the public by their role in stroke center designation, which is perhaps more meaningful to the public than the results of this measure. It would seem that hospitals could change their compliance with this measure quickly as evidenced by the rapid improvement. Agree with moderate ranking.

**The treatment is clearly beneficial - the value added by the measurement is a bit less persuasive. Assessing the cost benefit of the measure it perhaps beyond my knowledge, but as I see hospitals struggling to provide good care at the same time that they are trying to accurately and completely complete rapidly increasing record and accountability demands, the overall value is uncertain.

Criterion 5: Related and Competing Measures

Related or competing measures none

Related measures:

- CSTK-01 National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients;
- CSTK-02 Modified Rankin Score (mRS) at 90 Days;
- CSTK-03 Severity Measurement Performed for Subarachnoid Hemorrhage (SAH) and Intracerebral Hemorrhage (ICH) Patients (Overall Rate).

Harmonization - all measures are from the same developer and are harmonized

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Title: CSTK-06: Nimodipine Treatment Administered
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure title

Date of Submission: 1/15/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

Subcriterion 1a. Evidence to Support the Measure Focus
The measure focus is a health outcome or is evidence-based, demonstrated as follows:

- Health outcome:[3] a rationale supports the relationship of the health outcome to processes or structures of care.
- Intermediate clinical outcome, Process,[4] or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence[5] that the measure focus leads to a desired health outcome.
- Patient experience with care: evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information OR that patient experience with care is correlated with desired outcomes.
- Efficiency:[6] evidence for the quality component as noted above.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement.

5. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

6. Measures of efficiency combine the concepts of resource use and quality (NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1. This is a measure of:
Outcome

- ☐ Health outcome: Click here to name the health outcome
  Health outcome includes patient-reported outcomes (PRO, i.e., HRQoL/functional status, symptom/burden, experience with care, health-related behaviors)
- ☐ Intermediate clinical outcome: Click here to name the intermediate outcome

☐ Process: Nimodipine treatment within 24 hours of hospital arrival for subarachnoid hemorrhage to prevent vasospasm and improve neurological outcome

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME PERFORMANCE MEASURE *If not a health outcome, skip to 1a.3*

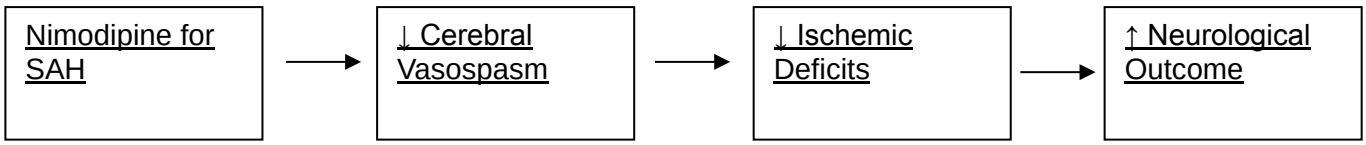
1a.2. Briefly state or diagram the linkage between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

Note: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the linkages between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



Nimodipine administration for SAH decreases cerebral vasopasm, thereby reducing the incidence and severity of ischemic deficits and improving neurological outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***
☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***
☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***
☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):
Connolly ES Jr, Rabinstein AA, Carhuapoma JR, Derdeyn CP, Dion J, Higashida RT, Hoh BL, Kirkness CJ, Naidech AM, Ogilvy CS, Patel AB, Thompson G, Vespa P, on belhalf of the American Heart Association Stroke Council, Council on Cardiovascular Radiology and Intervention, Council on Cardiovascular Nursing, Council on Cardiovascular Surgery and Anesthesia, Council on Clinical Cardiology. Guidelines for the management of aneurysmal subarachnoid hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2012;43:1-27.

URL:
http://stroke.ahajournals.org/content/early/2012/05/03/STR.0b013e3182587839.full.pdf

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Management of Cerebral Vasospasm and DCI After aSAH: Recommendations (Page 14)
Class I
1. Oral nimodipine should be administered to all patients with aSAH. (*Level of Evidence: A*).

(It should be noted that this agent has been shown to improve neurological outcomes but not cerebral vasospasm. The value of other calcium channel antagonists, whether administered orally or intravenously, remains uncertain.)

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Class I: Procedure/treatment SHOULD be performed/administered.

Level of Evidence A: Data derived from multiple randomized clinical trials or meta-analyses.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

ACCF/AHA Classification of Recommendation and Level of Evidence
Classification Types
Class II a: It is reasonable to perform procedure/administer treatment.
Class II b: Procedure/Treatment may be considered.
Class III: Procedure/Treatment not helpful/no proven benefit/may be harmful.

Level of Evidence
Level B: Data derived from a single randomized trial or nonrandomized studies.
Level C: Only consensus opinion of experts, case studies, or standard of care.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**
Same as 1a.4.1

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?
☐ Yes → ***complete section 1a.7***
☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION
1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE
1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Cochrane Systematic Review

Dorhout mees S, Rinkel GJE, Feigin VL, Algra A, van den Bergh WM, Vermeulen M, van Gijn J. Calcium antagonists for aneurismal subarachnoid haemorrhage. *Cochrane Database of Systematic Review* 2007, Issue 3. Art. No.: CD000277. DOI: 10.1002/14651858.CD000277.pub3.

http://web.a.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=f61f2e7c-5369-46a1-ab8f-3b02ca6fa8c7%40sessionmgr4005&vid=4&hid=4112

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**
The Cochrane Systematic Review identified quality of evidence based on risk of bias. System for determining risk of bias was explained in Chapter 8 of the Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.1, Table 8.4.a "A common classification scheme for bias" and Table 8.5.a "The Cochrane Collaboration's tool for assessing risk of bias", updated March 2011 http:// www.mrc-bsu.cam.ac.uk/cochrane/handbook.

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?
The information in the following questions in this section is based on the Cochrane Systematic Review cited in section 1a.6.

The systematic review addressed calcium antagonists for aneurysmal subarachnoid haemorrhage, comparing calcium antagonists with control, or a second calcium antagonist (magnesium sulphate) versus control in addition to another calcium antagonist (nimodipine) in both the intervention and control groups.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

An overall grade for the quality of quoted evidence was not assigned. In the systematic review, individual study quality was graded on a scale for risk of bias.

Based on reported adequacy of allocation of concealment and blinding, ten studies appear to be at low risk (Allen 1983; Ferro 1990; Haley 1993; Neil-Dwyer 1987; Ohman 1991; Petruk 1988; Philippon 1986; Pickard 1989; van den Bergh 2005; Wong 2006), two studies at high risk (Shibuya 1992; Veyna 2002), and four studies uncertain risk of bias (Han 1993; Luo 1996, Messeter 1987; Zhu 2001).

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.1, Section 8.5.3. page 8.12.

ALLOCATION CONCEALMENT Was allocation adequately concealed?

Criteria for a judgement of 'YES' (i.e. low risk of bias)
Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:
• Central allocation (including telephone, web-based and pharmacy-controlled randomization);
• Sequentially numbered drug containers of identical appearance;
• Sequentially numbered, opaque, sealed envelopes.

Criteria for the judgement of 'NO' (i.e. high risk of bias).
Participants or investigators enrolling participants could possibly foresee assignments and thus introduce selection bias, such as allocation based on:
• Using an open random allocation schedule (e.g. a list of random numbers);

• Assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or non-opaque or not sequentially numbered);
• Alternation or rotation;
• Date of birth;
• Case record number;
• Any other explicitly unconcealed procedure

Criteria for the judgement of 'UNCLEAR' (uncertain risk of bias).
Insufficient information to permit judgement of 'Yes' or 'No'. This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range
1983 - 2006

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)
16 randomized controlled trials.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Overall quality of evidence across 16 studies, involving 3361 patients, appears to be high. All randomized controlled trials in which any type of calcium antagonist was compared with control in patients with subarachnoid hemorrhage. The trials were comparable with respect to study design and selection of patients, and most has a well-balanced distribution of prognostic factors for poor outcome between treatment and control groups, which adds to the validity of results. Sensitivity analyses showed that the beneficial results in terms of risk reduction for poor outcome were robust in that the results did not appreciably change after exclusion of trials with questionable blinding or randomization of patients more than four days after SAH, or after inclusion of trials with very late enrollment or very early assessment.

This review found that the outcome after subarachnoid hemorrhage, in terms of survival and being independent in activities of daily living, is improved by treatment with calcium channel blockers (antagonists). If the largest trial (Haley 1993, n= 906) is excluded from the analysis, the results are no longer statistically significant, and therefore the evidence is not beyond all doubt. However, given the high likelihood of benefits and the modest risks associated with this treatment, the review authors concluded that calcium antagonists, in the form of oral nimodipine 60 mg every four hours, are useful in patients with subarachnoid hemorrhage from a ruptured aneurysm. Considering that most studies included in this review were performed prior to the availability of aneurysm coiling, further placebo-controlled trials of oral nimodiopine could provide a more definitive conclusion.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

The meta-analysis included 16 studies with a total of 3361 patients (1665 in the treatment group and 1696 in the control group). The effects of interventions were analyzed according to four outcomes:

Poor Outcome (death or dependence): 9 trials (n=2589) adequately reported functional outcome within 6 months after SAH. The relative risk (RR) for poor outcome (death or dependency) was 0.81 (95% confidence interval (CI) 0.72 to 0.92). the absolute risk reduction was 5.3%; the corresponding number of patients needed to treat to benefit (NNTB) to prevent a single poor outcome event was 19 (95% CI 1 to 51).

Case Fatality: 11 trials (n=2775); RR 0.87 (95% CI 0.73 to 1.02). There were no statistical differences between the subgroups of different calcium antagonists (P=1.0) or between the subgroups according to timing of outcome assessment (P=0.84).

Secondary Ischemia: Calcium antagonists versus placebo included 11 trials (n=2303). NNTB to prevent secondary ischemia was 7:1 and 9:1 to prevent CT-confirmed infarct. For oral nimodipine separately, RR for a clinical episode of secondary ischemia was 0.64 (95% CI 0.49 to 0.83), and for a CT-confirmed infarct 0.71 (95% CI 0.57 to 0.89). Analyses by other routes of administration of nimodipine showed no significant differences for the protective effect on secondary ischemia.

Rebleeding: Calcium antagonists versus placebo included 8 trials (n=2215). There was a statistically significant reduction in the frequency of rebleeding among patients allocated to calcium antagonist treatment: RR 0.75 (95% CI 0.57 to 0.98); absolute RR 3% corresponding to NNTB 39 (95% CI 21 to 431). There were no statistical differences between the subgroups of different calcium antagonists (P=0.72).

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?
This review indicates that nimodipine has neuroprotective properties and treatment can improve clinical outcomes for patients with subarachnoid hemorrhage (SAH). Secondary ischemia is a major complication of SAH and occurs in a third of all patients, most commonly within four to ten days after the hemorrhage, resulting in a poor outcome in half of the patients with this complication. Secondary ischemia is due to vasospasm which may result from an increase in calcium in the vascular smooth-muscle cells of cerebral arteries. Nimodipine counteracts the influx of calcium into the vascular smooth-muscle cell, thereby decreasing the rate of vasospasm; however, another effect of this calcium channel blocker is the induction of hypotension which can counteract potential benefits. This adverse event can generally be managed with oral administration of nimodipine and close patient monitoring. If administered intravenously, nimodipine can result in severe hypotension, cardiovascular collapse, and cardiac arrest.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
CSTK-06_Measure__Evidence_6.5.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Cerebral vasopasm is a serious complication following subarachnoid hemorrhage (SAH), occurring in 30% to 70% of patients and accounting for nearly 50% of the deaths in patients surviving to treatment (Bederson, 2009). Constriction of the arterial lumen results in diminished cerebral perfusion distal to the affected artery, which produces a delayed neurological deficit that may progress to cerebral infarction without early management of the ruptured aneurysm. The arterial narrowing that occurs in cerebral vasospasm is typically a transient or temporary event, lasting from a few days up to 3 weeks.
The main goal of current treatment is to prevent or limit the severity of cerebral vasospasm. Only two treatments are generally accepted as proven and valuable for the prevention of ischemic stroke and reduction of ischemic complications:

- Treatment with cerebroselective calcium channel blocker nimodipine-Nimotop (60mg po q4h for 21 days after hemorrhage or after hospital discharge if discharged within 21 days) (Leifer, 2011);
- Aggressive hypervolemic, hypertensive, hemodilution therapy (i.e., triple-H therapy) with pressor agents and volume expansion (colloids) while monitoring the central venous pressure (CVP) or pulmonary capillary wedge pressure (PCWP), following early clipping of the aneurysm.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Pilot Test Findings:
During the six-month pilot test (October 1, 2012-March 31, 2013), sixty-six sites submitted data for 10,218 completed patient records. For this measure, 1229 cases were assigned an ICD-9-CM Principal Diagnosis Code of 430 Subarachnoid Hemorrhage at discharge and captured in the denominator population, and 867 of these cases met the numerator requirements. The performance rates varied widely across sites for this measure with results ranging from a low of 0% to a high of 100%. The average rate for all sites collecting data for this measure was 71%, indicating a potential performance gap of approximately 30% if the optimal rate is 100%.

In January, 2015, The Joint Commission implemented data collection for the comprehensive stroke (CSTK) measure set to meet performance measurement requirements for its Comprehensive Stroke Certification Program. Below is the specified level of analysis for CSTK-06 Nimodipine Treatment Administered for the two quarters of data received to date for this measure.

1Q 2015: 572 denominator cases; 494 numerator cases; 39 hospitals; 0.86364 national aggregate rate; 0.85758 mean of hospital rates; 0.12332 standard deviation; 1.0 90th percentile rate; 0.97368 75th percentile rate/upper quartile; 0.86207 50th percentile rate/median rate; 0.75 25th percentile rate/lower quartile; and, 0.66667 10th percentile rate.

2Q 2015: 878 denominator cases; 711 numerator cases; 51 hospitals; 0.80978 national aggregate rate; 0.83311 mean of hospital rates; 0.17241 standard deviation; 1.0 90th percentile rate; 0.91892 75th percentile rate/upper quartile; 0.86667 50th percentile rate/median rate; 0.78947 25th percentile rate/lower quartile; and, 0.75 10th percentile rate.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
Citation:

Brain Aneurysm Foundation. Understanding: Brain Aneurysm Statistics and Facts.
http:// www.bafound.org/Statistics_and_Facts. Published 2014. Accessed June 5, 2014.

Summary:
Accurate early diagnosis is critical, as the initial hemorrhage may be fatal, may result in devastating neurologic outcomes, or may produce minor symptoms. Despite widespread neuroimaging availability, misdiagnosis or delays in diagnosis occurs in up to 25% of patients with subarachnoid hemorrhage (SAH) when initially presenting for medical treatment. Failure to do a scan results in 73% of these misdiagnoses. This makes SAH a low-frequency, high-risk disease.
Citation:
Mayer PL, Awad IA, Todor R, et. al. Misdiagnosis of symptomatic cerebral aneurysm: prevalence and correlation with outcome at four institutions. Stroke. 1996;27:1558-1563.

Summary:
Data analyzed: 1990's; 4 Connecticut, U.S. neurosurgical units; 217 SAH patients. Fifty-four (25%) of patients with subarachnoid hemorrhage initially received an incorrect diagnosis; most of them were in good clinical condition at presentation. The condition of the 54 patients with incorrect diagnosis worsened, usually as a result of recurrent bleeding, before definitive treatment was begun. Of the 163 patients given a correct diagnosis, the condition of only 2.5% worsened.

Citation:
Kowalski RG, Claassen J, Kreiter KT, Bates JE, Ostapkovich ND, Connolly ES, Mayer SA. Initial misdiagnosis and outcome after subarachnoid hemorrhage. JAMA. 2004 Feb 18;291(7): 866-869.

Summary:
Data analyzed: August 1996 – August 2001; U.S. tertiary care urban hospital; 482 SAH patients. Misdiagnosis of SAH occurred in 12% of patients and was associated with a smaller hemorrhage and normal mental status. Among individuals who initially present in good condition, misdiagnosis is associated with increased mortality and morbidity.

Citation:
Vermeulen MJ, Schull MJ. Missed diagnosis of subarachnoid hemorrhage in the emergency department. Stroke. 2007;38:1216-1221.

Summary:
Data analyzed: April 1, 2002 – March 31, 2005; 176 Canadian hospitals with emergency departments (EDs); 1603 patients hospitalized with a diagnosis of nontraumatic SAH. Among all nontraumatic SAH patients admitted to an Ontario hospitals, 5.4% had been misdiagnosed on a prior visit to an ED. About 1 in 20 persons with SAH are missed on their first presentation to an ED, and the risk is greater in patients with low acuity presentations.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This is the initial submission of this measure.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
Brain Aneurysm Foundation. Understanding: Brain Aneurysm Statistics and Facts.
http:// www.bafound.org/Statistics_and_Facts. Published 2014. Accessed June 5, 2014.

Ruptured brain aneurysms account for 3-5% of all new strokes. Women compared to men suffer from brain aneurysms at a ratio of 3:2.

Morgenstern LB, Smith MA, Lisabeth LD, Risser JM, Uchino K, Garcia N, Longwell PJ, McFarling DA, Akuwumi O, Al-Wabil A, Al-Senani F, Brown DL, Moyé LA. Excess Stroke in Mexican Americans compared with non-Hispanic Whites: the Brain Attack Surveillance in Corpus Christi Project. Am J Epidemiol. 2004 Aug 15;160(4):376-383.

The project studied 2,350 cerebrovascular events occurring from January 2000 to December 2002 in Nueces County, Texas. Of the completed strokes, 53% were in Mexican Americans. The crude cumulative incidence was 168/10,000 in Mexican Americans and

136/10,000 in non-Hispanic Whites. The subarachnoid risk age-adjusted risk ratio was 1.57 (95% confidence interval: 0.86, 2.89). Mexican Americans experienced a higher incidence of subarachnoid hemorrhage than non-Hispanic whites, adjusted for age.

_____

Schievink WI, Riedinger M, Jhutty TK, Simon P. Racial disparities in subarachnoid hemorrhage mortality: Los Angeles County, California, 1985-1998. Neuroepidemiology. 2004 Nov-Dec;23(6):299-305.

The number of SAH deaths was 2,897. The age-adjusted SAH mortality rate was 1.9 in whites, 2.7 in Hispanics, 3.0 in Asians and 3.7 in blacks. In those younger than 70 years of age, the SAH mortality rate among blacks was 2.2 times that of whites and 1.8 times that of Hispanics and Asians. The SAH mortality rate declined after age 70 in blacks. The SAH mortality rate was higher in women than in men in all races and it was highest in elderly Asian women (23.5 per 100,000). An inverse relationship was observed between income and SAH mortality rates in all racial groups except whites.

_____

Jaja BNR, Saposnik G, Nisenbaum R, Lo BWY, Schweizer TA, Thorpe KE, Macdonald RL. Racial/ethnic differences in outcomes following subarachnoid hemorrhage. 2013 Sep 10; DOI: 10.3171/2013.7.JNS13544.

Using 2005-2010 data from the Nationwide Inpatient Sample, Jaja and colleagues conducted a cross-sectional study of hospital discharges for patients whose principal diagnosis was SAH unrelated to trauma (n=31,631), In this study, inpatient mortality was the primary outcome and discharge to institutional care was the secondary outcome. The researchers found a crude inpatient mortality rate of 22% and a 42% rate of hospital discharge to institutional care. Multivariable analyses identified race/ethnicity as a significant predictor of both inpatient mortality (p=0.003) and discharge to institutional care (p < 0001). Hispanic, black, Native American, and Asian/Pacific Islander patients were compared to whites. The study found that Hispanic patients were the least likely to have a poor outcome, and Asian/Pacific Islander patients were most likely to have a poor outcome.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

A leading cause of morbidity/mortality, High resource use, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ˜ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. Of all strokes, 3% are subarachnoid hemorrhage (SAH) strokes. While SAH represents the smallest percentage of all strokes, it is the most deadly type of stroke resulting in more than a 50% mortality rate. Of the survivors, approximately half will suffer a permanent disability.

The mean lifetime cost of ischemic stroke, including inpatient care, rehabilitation, and follow-up as necessary for residual deficits are estimated at $140,048 per person. Death within 7 days, SAH, and stroke while hospitalized for another condition are associated with higher costs.

Nimodipine (Nimotop), a calcium channel blocker, is indicated to reduce the incidence and severity of ischemic deficits in patients with SAH from ruptured intracranial aneurysms. It improves neurological outcome by relaxing cerebral smooth muscle vasculature and preventing vasospasm.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Bayer Phamaceutical Corporation. Nimotop® (nimodipine) capsules US prescribing information [online]. http:// www.univgraph.com/bayer/inserts/nimotop.pdf. Published 2008. Accessed June 6, 2014.

Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Prevention, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.jointcommission.org/assets/1/6/CSTKManual2015August.pdf

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Copy_of_AppendixACSTKTables_ICD10codes-635878790579131970.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not Applicable.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
SAH patients for whom nimodipine treatment was administered within 24 hours of arrival at this hospital.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of Care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Six data elements are used to calculate the numerator. Data elements and definitions:
• Arrival Date - The earliest documented month, day, and year, the patient arrived at the hospital.

• Arrival Time - The earliest documented time (military time) the patient arrived at the hospital.

• Nimodipine Administration – Documentation that nimodipine was administered at this hospital. Nimodipine is a cerebroselective calcium channel blocker that inhibits calcium transport into vascular smooth muscle cells, thereby suppressing contractions. Nimodipine is used in the treatment of subarachnoid hemorrhage patients to prevent or limit the severity of cerebral vasospasm.
Allowable Values: Yes or No/UTD.

• Nimodipine Administration Date – The month, day, and year that the first dose of nimodipine was administered to a patient with subarachnoid hemorrhage at this hospital.

• Nimodipine Administration Time – The time (military time) for which the first dose of nimodipine was administered to a patient with subarachnoid hemorrhage at this hospital.

• Reason for Not Administering Nimodipine Treatment - Reasons for not administering nimodipine treatment:
o Nimodipine allergy
o Other reasons documented by physician/advanced practice nurse/physician assistant (physician/APN/PA) or pharmacist
Allowable Values: Yes or No/UTD.

Patients are eligible for the numerator population when the Nimodipine Administration Date and Nimodipine Administration Time minus the Arrival Date and Arrival Time are greater than or equal to zero minutes and less than or equal to 1440 minutes, OR the Reason for Not Administering Nimodipine Treatment equals allowable values 'Yes'.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
SAH patients

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Included Populations:
• Discharges with ICD-10-CM Principal Diagnosis Code for ischemic stroke as defined in Appendix A, Table 8.2a

7 data elements are used to calculate the denominator. Data elements and definitions:

• Admission Date: The month, day, and year of admission to acute inpatient care.

• Birthdate: The month, day, and year the patient was born.

• Clinical Trial: Documentation that during this hospital stay the patient was enrolled in a clinical trial in which patients with the same condition as the measure set were being studied. Allowable Values: Yes or No/UTD.

• Comfort Measures Only: Comfort Measures Only refers to medical treatment of a dying person where the natural dying process is permitted to occur while assuring maximum comfort. It includes attention to the psychological and spiritual needs of the patient and support for both the dying patient and the patient´s family. Comfort Measures Only is commonly referred to as "comfort care" by the general public. It is not equivalent to a physician order to withhold emergency resuscitative measures such as Do Not Resuscitate (DNR).
Allowable Values: 1 (Day 0 or 1); 2 (Day 2 or after); 3 (Timing unclear); 4 (Not documented/UTD).

• Discharge Date - The month day and year the patient was discharged from acute care, left against medical advice or expired during the stay.

• Discharge Time – The documented time (military time) the patient was discharged from acute care, left against medical advice or expired during the stay.

• ICD-10-CM Principal Diagnosis Code: The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
• Patients less than 18 years of age

• Patients who have a Length of Stay greater than 120 days
• Patients with Comfort Measures Only documented on the day of or day after hospital arrival
• Patients enrolled in Clinical Trials
• Patients discharged within 24 hours of arrival at this hospital

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
• Patients less than 18 years of age.
o Patient age (in years) equals Admission Date minus Birthdate.

• Patients who have a Length of Stay greater than 120 days.
o Length of Stay (in days) equals Discharge Date minus Admission Date.

• Patients with Comfort Measures Only documented:
o Physician/APN/PA documentation of comfort measures only (hospice, comfort care, etc.) when the earliest day of documented CMO was on the day of arrival (Day 0) or Day after arrival (Day 1).

• Patients enrolled in a Clinical Trial.
o Patients are excluded if "Yes" is selected for Clinical Trial.

• Patients who expire within 24 hours of arrival at this hospital
o Patients expiration equals Discharge Date and Discharge Time minus Arrival Date and Arrival Time greater than or equal to 0 minutes and less than 1440 minutes

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not Applicable

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not Applicable

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not Applicable

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk*

adjustment; etc.)
Comprehensive Stroke (CSTK) Initial Patient Population Algorithm

Variable Key: Patient Age, Initial Patient Population Reject Case Flag, Length of Stay, Sub-Population 1 Flag, Sub-Population 2 Flag, and Sub-Population 3 Flag.

1. Start CSTK Initial Patient Population logic sub-routine. Process all cases that have successfully reached the point in the Transmission Data Processing Flow: Clinical which calls this Initial Patient Population Algorithm. Do not process cases that have been rejected before this point in the Transmission Data Processing Flow: Clinical.

2. Check ICD-10-CM Principal Diagnosis Code
a. If the ICD-10-CM Principal Diagnosis Code is not on Table 8.1 and 8.2, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the ICD-10-CM Principal Diagnosis Code is on Table 8.1 or 8.2, continue processing and proceed to the Patient Age calculation.

3. Calculate Patient Age. Patient Age, in years, is equal to the Admission Date minus the Birthdate. Use the month and day portion of admission date and birthdate to yield the most accurate age.

4. Check Patient Age
a. If the Patient Age is less than 18 years, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the Patient Age is greater than or equal to 18 years, continue processing and proceed to Length of Stay Calculation.

5. Calculate the Length of Stay. Length of Stay, in days, is equal to the Discharge Date minus the Admission Date.

6. Check Length of Stay
a. If the Length of Stay is greater than 120 days, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the Length of Stay is less than or equal to 120 days, the patient is in the CSTK Initial Patient Population.

7. Set the Initial Patient Population Reject Case Flag to equal No. Continue processing and proceed to the ICD-10-CM Principal Diagnosis Code to determine the CSTK sub-population.

8. Initialize Sub-Population 1 Flag, Sub-Population 2 Flag and Sub-Population 3 Flag to No.

9. Check ICD-10-CM Principal Diagnosis Code
a. If the ICD-10-CM Principal Diagnosis Code is on 8.2, the patient is in the CSTK Sub-Population 3 and is eligible to be sampled for the CSTK Sub-Population 3. Set Sub-Population 3 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the ICD-10-CM Principal Diagnosis Code is on 8.1, continue processing and proceed to ICD-10-PCS Principal Or Other Procedure Codes.
i. If at least one ICD-10-PCS Principal or Other Procedure Codes is on Table 8.1a or 8.1b, the patient is in the CSTK Sub-Population 2 and is eligible to be sampled for the CSTK Sub-Population 2. Set Sub-Population 2 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
ii. If none of the ICD-10-PCS Principal Or Other Procedure Codes are on Table 8.1a or 8.1b, the patient is in the CSTK Sub-Population 1 and is eligible to be sampled for the CSTK Sub-Population 1. Set Sub-Population 1 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

CSTK-06: Nimodipine Treatment Administered

Numerator: SAH patients for whom nimodipine treatment was administered within 24 hours of arrival at this hospital.

Denominator: SAH patients

Variable Key: Timing I, Timing II
1. Start processing. Run cases that are included in the Comprehensive Stroke (CSTK) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check ICD-10-CM Principal Diagnosis Code
a. If ICD-10-CM Principal Diagnosis Code is not on Table 8.2a, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
b. If ICD-10-CM Principal Diagnosis Code is on Table 8.2a, continue processing and proceed to Comfort Measures Only.

3. Check Comfort Measures Only
a. If Comfort Measures Only is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Comfort Measures Only equals 1, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If Comfort Measures Only equals 2, 3, or 4, continue processing and proceed to Clinical Trial.

4. Check Clinical Trial
a. If Clinical Trial is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Clinical Trial equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If Clinical Trial equals No, continue processing and proceed to Arrival Date.

5. Check Arrival Date
a. If Arrival Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Arrival Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Arrival Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to Arrival Time.

6. Check Arrival Time
a. If Arrival Time is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Arrival Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Arrival Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to Discharge Date.

7. Check Discharge Date
a. If Discharge Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Discharge Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Discharge Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to the Discharge Time.

8. Check Discharge Time
a. If Discharge Time is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Discharge Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Discharge Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to the Timing I calculation.

9. Calculate Timing I. Timing I, in minutes, is equal to the Discharge Date and the Discharge Time minus the Arrival Date and Arrival Time.
a. If the time in minutes is greater than or equal to zero and less than 1440, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
b. If the time in minutes is greater than or equal to 1440, the case will proceed to Nimodipine Administration.

10. Check Nimodipine Administration
a. If Nimodipine Administration is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Nimodipine Administration equals No, continue processing and proceed to step 14 and check Reason for Not Administering

Nimodipine Treatment.
c. If Nimodipine Administration equals Yes, continue processing and proceed to Nimodipine Administration Date.

11. Check Nimodipine Administration Date
a. If Nimodipine Administration Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Nimodipine Administration Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Nimodipine Administration Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to Nimodipine Administration Time.

12. Check Nimodipine Administration Time
a. If Nimodipine Administration Time is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Nimodipine Administration Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Nimodipine Administration Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to the Timing II calculation.

13. Calculate Timing II. Timing II, in minutes, is equal to the Nimodipine Administration Date and the Nimodipine Administration Time minus the Arrival Date and Arrival Time.
a. If the time in minutes is greater than 1440, the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
b. If the time in minutes is greater than or equal to zero and less than or equal to 1440, the case will proceed to a Measure Category Assignment of E and will be in the numerator population. Stop processing.

14. Check Reason for Not Administering Nimodipine Treatment
a. If Reason for Not Administering Nimodipine Treatment is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Reason for Not Administering Nimodipine Treatment equals No, the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Reason for Not Administering Nimodipine Treatment equals Yes, the case will proceed to a Measure Category Assignment of E and will be in the numerator population. Stop processing.

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter for the measure set cannot sample. Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

The sub-population for the CSTK-06 measure Initial Patient Population is CSTK 3-Hemorrhagic Stroke. The CSTK 3 sub-population sampling group includes patients admitted to the hospital for inpatient acute care if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.2, a Patient Age (Admission Date – Birthdate) > 18 years, and a Length of Stay (Discharge Date - Admission Date) less than or equal to 120 days.

Quarterly Sampling
The Quarterly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for CSTK sub-population 3. Hospitals performing quarterly sampling for the CSTK-06 measure must ensure that its Initial Patient Population and sample size meet the following conditions for the CSTK 3 sub-population sampling group:

Quarterly Sample Size Based on CSTK Sub-population 3 for Hemorrhagic Stroke (Table 3)

Sub-Population 3: If "N" > 750, then 'n' 150
Minimum Required Sample Size: 150 records

Sub-Population 3: If "N" 376-750, then 'n' 20%
Minimum Required Sample Size: 20% of Sub-Population records

Sub-Population 3: If "N" 76-375, then 'n' 75
Minimum Required Sample Size: 75 records

Sub-Population 3: If "N" < 75, then 'n' 100%
Minimum Required Sample Size: No sampling; 100% Sub-Population records required

Monthly Sampling
The Monthly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for CSTK sub-population 3. Hospitals performing monthly sampling for the CSTK-06 measure must ensure that its Initial Patient Population and sample size meet the following conditions for CSTK 3 sub-population sampling group:

Monthly Sample Size Based on CSTK Sub-population 3 for Hemorrhagic Stroke (Table 6)

Sub-Population 3: If "N" > 250, then 'n' 50
Minimum Required Sample Size: 50 records

Sub-Population 3: If "N" 126-250, then 'n' 20%
Minimum Required Sample Size: 20% of Sub-Population size records

Sub-Population 3: If "N" 26-125, then 'n' 25
Minimum Required Sample Size: 25 records

Sub-Population 3: If "N" < 25, then 'n' 100%
Minimum Required Sample Size: No sampling; 100% Sub-Population records required

S.21. Survey/Patient-reported data *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not Applicable

S.22. Missing data (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

S.23. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Paper Medical Records

S.24. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
A web-based data collection tool was developed by The Joint Commission for the pilot test process. Currently, hospitals have the flexibility of creating their own tool modeled after the pilot tool or they may develop their own data collection tools using the data element dictionary and allowable values specified in the implementation guide. Hospitals also have the option of selecting a vendor-developed data collection tool which has been verified by The Joint Commission.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not Applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2863_MeasureTesting_MSF6.5.docx

Measure Title: CSTK-06: Nimodipine Treatment Administered

Date of Submission: 1/15/2016

Type of Measure: Process

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions.
15. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.
16. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Not Applicable

1.3. What are the dates of the data used in testing? October 1, 2012 –March 31, 2013; first and second quarter 2015.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Hospitals were recruited for the pilot test of the measures via an open call on The Joint Commission web site. An announcement of the call with a link to The Joint Commission web site was also posted on the American Heart Association/American Stroke Association web site. Hospitals were selected with the intent to capture variability related to ownership, size, type of facility and location. Eighty-two hospitals from twenty-seven states were selected from more than 120 volunteers to participate in the six-month pilot test of the measures. Twenty hospitals withdrew during the pilot test citing lack of resources to complete the project. Sixty-two hospitals submitted data for each month of the six-month pilot test. An additional four hospitals submitted data for one or more months.

Sixty-six hospitals contributed data for the analysis of the measures:

Ownership:

For Profit	14
Not for Profit	52

Bed Size:	
Less than 100	0
100 – 199	3
200 – 299	7
300 – 499	26
Greater than 500+	30

Located in 27 states:
Alabama
Arizona
California
Colorado
Florida
Georgia
Illinois
Indiana
Kentucky
Louisiana
Maryland
Massachusetts
Michigan
Minnesota
Missouri
Nevada
New Jersey
New York
North Carolina
Ohio
Oregon
Pennsylvania
Tennessee
Texas
West Virginia
Washington
Wisconsin

Other:	
Teaching	41
Non-teaching	25

Urban	59
Rural	7

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

During the six-month pilot test, sixty-six hospitals submitted data for 1,229 inpatient records. The cases included patients greater than 18 years of age, male and female, all races, and all payers. Age, gender, racial, and payer distribution are not known because the results were de-identified. The patient population was comprised of all hemorrhagic stroke patients discharged with an ICD-9-CM Principal Diagnosis Code for subarachnoid hemorrhage (SAH).

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Data from the six month pilot test of the measures were used for reliability testing and face validity. To test the empirical validity of the measures, two quarters of data from 42 Joint Commission certified comprehensive stroke centers were used to conduct a secondary analysis.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Reliability testing was performed at twelve participating pilot sites. Testing was conducted on a stratified random sample of records selected from each organization at the organization and measure category level. Hospitals were visited by teams of two Joint Commission staff during April, May, June, July, and August 2013. During these visits, Joint Commission staff re-abstracted records previously abstracted by hospital staff. A total of 281 records were re-abstracted. In cases of disagreement between the re-abstracted and original abstraction on a data element, reasons for disagreement were determined and adjudication was made as to whether original or re-abstraction findings were correct. Reliability was addressed by comparing the original abstracted and adjudicated re-abstracted values, with the adjudicated value serving as the gold standard. The data analysis included both the percent agreement and the kappa statistic to adjust for chance agreement for categorical data elements.

2a2.3. For each level checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data Element	Number of Mismatches	Match Rate	Kappa
Arrival Date	7	97.%%	NA
Arrival Time	50	82.2%	NA
Discharge Date	9	96.8%	NA
Discharge Time	80	71.5%	NA
Nimodipine Administration	4	98.6%	0.93
Nimodipine Administration Date	7	97.5%	NA
Nimodipine Administration Time	15	95.7%	NA
Reason for Not Administering Nimodipine Treatment	16	94.3%	-0.0274

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

A statistical measure of inter-rater reliability is Cohen's Kappa, which ranges generally from 0.0 to 1.0 (although negative numbers are possible), where large values mean better reliability and values near zero suggest that agreement is attributable to chance alone. It indicates the proportion of agreement not expected by chance alone (e.g., Kappa of 0.6 means that raters agreed 60% of the time over and above what would be expected by chance alone).

Landis & Koch, 1977 offers the following clarification of Kappa interpretation:

< 0	Poor agreement
0.00-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-1.00	Almost perfect agreement

Other authors (Cicchetti & Sparrow; Fleiss) have suggested additional classifications for interpreting the Kappa statistic, but all seem to indicate Kappa > 0.60 is desirable.

The statistical measure of inter-rater reliability is suitable for the categorized data type, therefore the measure is applied to the following data elements: *Nimodipine Administration* and *Reason for Not Administering Nimodipine Treatment*. The Kappa values indicate almost perfect agreement for the data element *Nimodipine Administration*; the Kappa value indicates poor agreement for the data element *Reason for Not Administering Nimodipine Treatment;* this was due to the low marginal rate of 6% for those with a documented reason and the lack of agreement for these on whether or not there was a valid reason. Except for this data element, we believe the results demonstrate acceptable reliability of the assessment data used in the performance measure.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

X **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

X **Empirical validity testing**

X **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Face Validity:

Measure face validity was assessed via survey and focus groups of hospitals participating in the pilot test. Focus group discussions were held at all test sites visited, during which we received feedback as to whether the measure, data elements, and definitions accurately reflected existing evidence. All of the respondents indicated that all aspects of the measures accurately reflected current evidence. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant.

To determine feasibility and identify areas for potential revision, test sites were asked to electronically rate the clarity of numerator statements, denominator statements, and measure information forms (MIFs) on a five point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree). Data elements and associated tables were evaluated for clarity, accuracy, data availability and accessibility.

Empirical Validity:

Measure convergent validity was assessed using hospitals patient level data from The Joint Commission warehouse. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous face validity testing. We conducted a secondary analysis to help interpret results; correlation among CSTK process measures.

The data were comprised of first and second quarter 2015 submissions. This included 42 hospitals submitting 65,389 inpatient records for the selected CSTK measures CSTK-1, CSTK-2, CSTK-3 and CSTK-6. The hospital's selection was based on those hospitals that reported 6 months of data and had 30 or more denominator cases for the period.

Comprehensive Stroke (CSTK) Initial Patient Population
The CSTK Initial Patient Population is unique in that it is comprised of three distinct subpopulations: ischemic stroke patients who do not undergo a reperfusion therapy (i.e., procedure), ischemic stroke patients who undergo a reperfusion therapy (IV t-PA, IA t-PA, or mechanical endovascular reperfusion (MER) therapy), and hemorrhagic stroke patients.

Subpopulation 1: Ischemic Stroke Without Procedure
This subpopulation comprises ischemic stroke patients who are admitted to the hospital for inpatient acute care and do not undergo a reperfusion procedure (CSTK-01 measure). Patients are included in the CSTK-1 Ischemic Stroke Without Procedure subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.1, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

Subpopulation 2: Ischemic Stroke With IV t-PA, IA t-PA, or MER
This subpopulation comprises ischemic stroke patients who receive IV t-PA, IA t-PA, or MER procedures during the hospital stay (CSTK-01 and CSTK-02 measures). Patients admitted to the hospital for inpatient acute care are included in the CSTK-2 Ischemic Stroke With IV t-PA, IA t-PA, or MER subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.1 and ICD-10-PCS Principal or Other Procedure Codes as defined in Appendix A, Table 8.1a OR Table 8.1b, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

Subpopulation 3: Hemorrhagic Stroke
This subpopulation comprises hemorrhagic stroke patients admitted to the hospital for inpatient acute care (CSTK-03 and CSTK-06 measures). Patients are included in the CSTK-3 -Hemorrhagic Stroke subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.2, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Pilot Test Data:
Descriptive statistics for the measure: N= 66 hospitals
Overall rate=60% (SD=31%), min=0%, max=100%

Additional analyses:
CSTK03 and CSTK06 are measures for hemorrhagic stroke. The measure rate among these measures are expected to be correlated. The Pearson correlation coefficient is calculated.

Pearson Correlation Coefficient interpretation: the range is from -1 to 1. When the correlation coefficient is close to +1 or -1, it means that there is strong correlation; p value is utilized to determine if the corrlation coefficient is significant or not. If it is less than 0.05, then the conclusion usually is significant.

The table below demonstrates that the CSTK06 rate was negtively and slightly correlated to the CSTK03 measure. Because the data is not normally distributed, the correlation coefficient on correlation may not be the best measure of association.

Pearson Correlation Coefficients, N = 66 Prob > |r| under H0: Rho=0

	CSTK03	CSTK06
Correlation Coefficient	-0.02229	1.00000
P value	0.8590	

The average rating for measure CSTK-06 numerator and denominator statements, including the clarity of numerator and denominator inclusions and exclusions was 4.16, indicating that the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

Data element validity was evaluated for clarity and accuracy, as well as, data availability and accessibility. The percentage of agreement for new data elements developed specifically for CSTK-06 and other measures in the CSTK measure set is detailed in the table below:

Data Element Name	Clarity / Accuracy	Availability / Accessibility
Discharge Time	95.45%	86.36%
Nimodipine Administration	97.67%	86.36%
Nimodipine Administration Date	100.00%	86.36%
Nimodipine Administration Time	100.00%	86.36%
Reason for Not Administering Nimodipine Treatment	95.45%	86.36%

1Q and 2Q 2015 Data:
Overall descriptive statistics for CSTK selected measures: N = 42 certified comprehensive stroke hospitals: n = 65,389

CSTK-01
Median: 89%
Percentile 10%: 70%
Percentile 25%: 80%
Percentile 75%: 93%
Percentile 90%: 95%

CSTK-02
Median: 95%
Percentile 10%: 63%
Percentile 25%: 75%
Percentile 75%: 100%
Percentile 90%: 100%

CSTK-03
Median: 61%
Percentile 10%: 33%
Percentile 25%: 48%
Percentile 75%: 79%
Percentile 90%: 89%

CSTK-06
Median: 86%
Percentile 10%: 74%
Percentile 25%: 80%
Percentile 75%: 93%
Percentile 90%: 100%

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
CSTK_1	42	0.83687	0.17097	35.14857	0.00267	1.00000
CSTK_2	41	0.85894	0.20516	35.21648	0.20000	1.00000
CSTK_3	42	0.59308	0.22979	24.90933	0.07273	0.92969
CSTK_6	42	0.83555	0.17123	35.09314	0	1.00000

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations				
	CSTK_1	CSTK_2	CSTK_3	CSTK_6
CSTK_1	1.00000 42	0.41205 0.0074 41	0.65340 <.0001 42	0.69186 <.0001 42
CSTK_2	0.41205 0.0074 41	1.00000 41	0.28658 0.0693 41	0.31724 0.0433 41
CSTK_3	0.65340 <.0001 42	0.28658 0.0693 41	1.00000 42	0.55910 0.0001 42
CSTK_6	0.69186 <.0001 42	0.31724 0.0433 41	0.55910 0.0001 42	1.00000 42

The CSTK measures table shows a positive correlation and statistical significance which indicates that hospitals with high quality on one CSTK measure tend to have high correlations on the other stroke measures.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)
2863

1Q and 2Q 2015 Data:
Overall the positive inter-correlations indicates convergent validity of all the measures.
They are positively correlated with other evidence-based processes of care.

2b3. EXCLUSIONS ANALYSIS .
NA ☐ no exclusions — *skip to section 2b4*
2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Pilot Test Data:
The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age

- Patients who have a Length of Stay greater than 120 days
- Patients with Comfort Measures Only documented on day of or day after hospital arrival
- Patients enrolled in a Clinical Trial
- Patients discharged within 24 hours of arrival at this hospital

1Q and 2Q 2015 Data:
The following exclusions were analyzed by subpopulation and measure for frequency and variability across providers:

- Patients less than 18 years of age
- Patients who have a Length of Stay greater than or equal to 120 days
- Patients with *Comfort Measures Only* documented on the day of or day after hospital arrival
- Patients enrolled in a Clinical Trial
- Patients discharged within 24 hours of arrival at this hospital

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Pilot Test Data:
There were 10668 admissions included in the initial cohort and diagnosed hemorrhagic stroke. From among the 10668 admissions in 66 hospitals, the descriptive statistics are given below.

Exclusion: Patient not in CSTK Initial Patient Population
Note: A case was excluded from the Initial Patient Population as determined by the following:

- Patients less than 18 years of age: Overall Occurrence n = 39 (0.37%)
- Patients who have a Length of Stay greater than 120 days: Overall Occurrence n = 12 (0.11%)

There were 3830 admissions included in the initial cohort and diagnosed hemorrhagic stroke. From among the 3830 admissions in 66 hospitals, the descriptive statistics are given below.

Exclusion: Clinical Trial
Overall Occurrence n = 5
Overall Occurrence Percentage: 0.13%

Exclusion: Comfort Measures
Overall Occurrence n = 105
Overall Occurrence Percentage: 2.74%

Exclusion: Patients discharged within 24 hours of arrival at this hospital
Overall Occurrence n = 26
Overall Occurrence Percentage: 0.68%

1Q and 2Q 2015 Data:
There were 65,389 admissions selected from the initial cohort. From among the 65,389 admissions in 42 hospitals, the descriptive statistics are given below.

<u>Applied To Measure CSTK-01 CSTK-03 CSTK-06</u>
Exclusion: Comfort Measures - 1 Day 0 or 1:
Overall Occurrence n = 1,300
Overall Occurrence Percentage: 2%
Minimum: 0.47%

Median: 3%
Maximum: 6%

Applied To Measure CSTK-06
Exclusion: Patients enrolled in clinical trials
Overall Occurrence n = 26
Overall Occurrence Percentage: 0.35%
Minimum 0.2%
Median: 0.70%
Maximum: 1.2%

Applied To Measure CSTK-6
Exclusion: Patients discharged within 24 hours of arrival at this hospital
Overall Occurrence n = 20,690
Overall Occurrence Percentage: 64%
Minimum: 48%
Median: 55%
Maximum: 100%
NMISS= 20,232 missing observation

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Pilot Test Data:
According to the overall occurrences in 2b3.2, the overall frequency of exclusions is low for those in the measure denominator. The distribution of exclusions across hospitals is very narrow indicating that the occurrence is random and likely would not bias performance results.

It is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a Length of Stay greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Patients with Comfort Measures Only documented
Rationale: It is inappropriate to treat such patients beyond those measures necessary for comfort.

Patients enrolled in a Clinical Trial
Rationale: Cases excluded to prevent potential conflicts with trial protocols.

Patients discharged within 24 hours of arrival at this hospital
Rationale: It is inappropriate to include patients whose hospital stay is less than the timeframe for administration of the first dose of nimodipine.

1Q and 2Q 2015 Data:
The overall frequency of exclusions is low for those in the measure denominator. The distribution of exclusions across hospitals is narrow indicating that the occurrence is random and likely would not bias performance results.

It is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a *Length of Stay* greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Patients with *Comfort Measures Only* documented
Rationale: It is inappropriate to treat such patients beyond those measures necessary for comfort.

Patients enrolled in a Clinical Trial
Rationale: Nimodipine therapy as specified by this measure may compromise a clinical trial.

Patients discharged within 24 hours of arrival at this hospital
Rationale: It is inappropriate to include patients whose hospital stay is less than the timeframe for administration of the first dose of nimodipine.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

Not Applicable

2b4.1. What method of controlling for differences in case mix is used?
X **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
Not Applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care and not related to disparities*)
Not Applicable

2b4.4. What were the statistical results of the analyses used to select risk factors?
Not Applicable

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)
Not Applicable

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
if stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
Not Applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Not Applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Not Applicable

2b4.9. Results of Risk Stratification Analysis:
Not Applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)
Not Applicable

***2b4.11. Optional Additional Testing for Risk Adjustment** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*
Descriptive statistics for the performance measure scores for all tested entities were constructed. These statistics were the mean, standard deviation, median, minimum, and maximum scores.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)
A meaningful difference was defined as a spread of more than 30 percentage points between the minimum and maximum scores or between the median and maximum pilot hospital rates.

Results for CSTK-06 were:
Descriptive statistics for measure: N=66 hospitals
Overall rate=60.1%
Standard Deviation=31.1%
Minimum=0%
Median=0.70%
Maximum=100%

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The results are interpreted as showing a meaningful spread between both the medium and maximum scores and between the minimum and maximum scores. There is evidence of considerable room for improvement of performance.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Not Applicable

Note: *This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different datasources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not Applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not Applicable

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not Applicable

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other

If other: Data element allowable values are selected, either manually or electronically, from clinical and coded data available in medical record documentation. All medical record documentation is used in the abstraction process. Vendor data collection tools are used to import data elements needed for measure rate calculation.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract this measure electronically, so offers a chart-abstracted version which allows for data capture from unstructured data fields. The Joint Commission plans to retool the measure for capture from electronic sources within the next several years.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

This measure was collected during a 6 month pilot process for Joint Commission Comprehensive Stroke Certification. It was learned via discussions and written pilot evaluation that this measure did not require revision and could be retained as written.

Other information impacting the feasibility and implementation of the measure was also obtained from the pilot process and is summarized as follows:

Staff Training and Education:

To prepare for and support continuous data collection throughout the pilot test, a total of ten hours were spent on staff training and education. Training was accomplished via two 2-hour webinars and monthly conference calls with pilot site participants.

Case Identification/Medical Record Retrieval:
Case identification was not a problem; cases were identified by the ICD-9-CM principal diagnosis codes for subarachnoid hemorrhage (430). Record retrieval time varied depending on the type of medical record. On average, 10 minutes were spent for record retrieval with more time spent to retrieve a paper record than electronic health record.

Case Selection:
For the pilot test of the measures, 100% record review without sampling was requested. During the pilot process it was noted that some facilities treated more than 25 hemorrhagic stroke patients per month. A sampling methodology for the hemorrhagic sub-population was added to the measure specifications post-pilot to ease abstraction burden for hospitals with a large number of hemorrhagic cases.

Data Abstraction:
Medical record review to abstract all data elements required for the measure set averaged 45 minutes per record. Time spent on record review varied with case complexity and the number of procedures and interventions performed, as well as, the number of data elements collected for the measure. In general, hemorrhagic stroke cases required less time for review than did ischemic stroke cases.

Data abstraction was primarily done by nurses, (e.g., Registered Nurse(s) with a quality improvement background, Stroke Coordinators, and Advanced Practice Nurses). Some pilot sites reported that the abstractor reviewed the record with the medical director or neurologist at least initially to identify documentation of measure specific data elements. Data specialists or administrative staff were utilized to enter abstracted data into the on-line data collection tool.

Cost of Data Abstraction:
Using 2012 national wage averages, it is estimated that the cost per case to abstract for this measure was approximately $3.50.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.
This measure is one in a set of measures implemented with discharges effective January 1, 2015 for Joint Commission Comprehensive Stroke Certification. There are no fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Public Health/Disease Surveillance Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Quality Improvement (Internal to the specific organization) Disease-Specific Care Certification for Comprehensive Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx

4a.1. **For each CURRENT use, checked above, provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

• Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Not Applicable

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Not Applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Sizeable improvement has been noted since the pilot test of the measure based on the first two quarters of Joint Commission ORYX performance measurement data. The initial performance measure gap of approximately 30% has decreased by 10% as demonstrated by a national aggregate rate of 81% (N=51); however, a 25% gap exists for the lowest decile of hospitals.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not Applicable

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no unintended negative consequences reported or detected during testing or since implementation of the measure specifications.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Not applicable

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Not applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Karen, Kolbusz, kkolbusz@jointcommission.org, 630-792-5931-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Ann, Watt, awatt@jointcommission.org, 630-792-5944-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel was to provide advisory oversight in the literature review, measure construct and content, review of testing results, and endorsement of draft and finalized measures. Additionally they may be called upon in the future to provide measure content oversight and updates.

Technical Advisory Panel

Harold P. Adams, Jr., MD
Professor of Neurology
Director of the Division of Cerebrovascular Diseases
Director of the Stroke Center
University of Iowa Health Care
Iowa City, IA

Mark J. Alberts, MD
Clinical Vice-Chair
Department of Neurology and Neurotherapeutics
The University of Texas Southwestern Medical Center
Dallas, TX

Anne W. Alexandrov, PhD, RN, FAAN
Health Outcomes Institute
Fountain Hills, AZ

Andrew W. Asimos, MD
Director of Emergency Stroke Care
Carolinas Medical Center
Charlotte, NC

Thomas P. Bleck, MD, FCCM
Professor of Neurological Science, Neurosurgery, Internal Medicine, and Anesthesiology
Associate Chief Medical Officer, Critical Care
Rush Medical College
Chicago, IL

Darlene Bradley, RN, PhDc, MSN, MAOM

Director Emergency/Trauma Services
University of California Irvine Medical Center
Orange, CA

Mary G. George, MD, MSPH FACAS
Medical Officer
Division for Heart Disease and Stroke Prevention
Centers for Disease Control and Prevention
Atlanta, GA

Martin Gizzi, MD, PhD
Legacy Health System
Portland, OR

James, Grotta, MD
Professor and Chairman, Department of Neurology
Chief, Vascular Neurology
University of Texas Medical School-Houston
Houston, TX

Edward C. Jauch, MD, MS
Professor, Division of Emergency Medicine and Department of Neurosciences
Medical University of South Carolina
Charleston, SC

Irene Katzan, MD, MS
Director, Neurological Institute Center for Outcomes Research and Evaluation
Cleveland Clinic
Cleveland, OH

Jawad F. Kirmani, MD
Director of Stroke and Cerebrovascular Program
JKF Medical Center
Edison, NJ

Brajesh K. Lal, MD
University of Maryland Medical Center,
Division of Vascular Surgery
Baltimore, MD

Kimberly E. Levasseur-Franklin, PharmD, BCPS
Critical Care Pharmacist, Neuro/Spine ICU
Clinical Leader, ICU Pharmacist Work Group
Northwestern Memorial Hospital
Chicago, IL

Barbara Lutz, PhD, RN, CRRN, APHN-BC, FAHA, FNAP, FAAN
McNeill Distinguished Professor
School of Nursing / College of Health and Human Services
University of North Carolina-Wilmington
Wilmington, NC

Cameron McDougall, MD, FRCSC
Barrow Neurological Institute
Phoenix, AZ

Kathy Morrison, RN, MSN, CNRN

Stroke Program Coordinator
Penn State Hershey Medical Center
Hershey, PA

Charles J. Prestigiacomo, MD, FAANS, FACS
Professor, Department of Neurological Surgery, Radiology, Neurology and Neurosciences
Director, Residency Program in Neurological Surgery
Director, Cerebrovascular and Endovascular Surgery
University of Medicine and Dentistry of New Jersey
Newark, NJ

Howard A. Rowley, MD
Chief of Neuroradiology
Joseph Sackett Professor of Radiology
Professor of Radiology, Neurology, and Neurosurgery
University of Wisconsin
Madison, WI

Jeffrey Saver, MD, FAHA, FAAN
Professor of Neurology
Stroke Center Director
Professor of Medicine
David Geffen School of Medicine at UCLA,
Ronald Reagan UCLA Medical Center
Los Angeles, CA

Lee H. Schwamm, MD, FAHA
Vice Chair of Neurology, Professor of Neurology
Harvard Medical School
Chair, American Heart Association GWTG Steering Committee
Director, Stroke Program/Telemedicine
Massachusetts General Hospital
Boston, MA

Deborah Summers, MSN, ACNS-BC, CNRN, CCRN, FAHA
Stroke Program Coordinator
St. Luke's Brain and Stroke Institute
Kansas City, MO

Dileep R. Yavagal, MD
Director, Interventional Neurology
Assistant Professor, Clinical Neurology and Neurosurgery
University of Miami and Jackson Memorial Hospital
Miami, FL

Osama O. Zaidat, MD, MS, FAAN, FAHA
St. Vincent Mercy Hospital
Toledo, OH

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2015

Ad.3 **Month and Year of most recent revision:** 08, 2015

Ad.4 **What is your frequency for review/update of this measure?** biannual

Ad.5 **When is the next scheduled review/update for this measure?** 02, 2016

that, and 2) users participating in Joint Commission accreditation, including ORYX® vendors, are required to update their software and associated documentation based on the published manual production timelines. Ad.7 **Disclaimers:**
Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2864
Measure Title: CSTK-01: National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients
Measure Steward: The Joint Commission
Brief Description of Measure: Proportion of ischemic stroke patients age 18 years or older for whom an initial NIHSS score is performed prior to any acute recanalization therapy (i.e., intra-venous (IV) thrombolytic (t-PA) therapy, or intra-arterial (IA) thrombolytic (t-PA) therapy, or mechanical endovascular reperfusion (MER) therapy) in patients undergoing recanalization therapy and documented in the medical record, or documented within 12 hours of arrival at the hospital emergency department in patients who do not undergo recanalization therapy.

This is the first in a set of measures developed for Joint Commission Comprehensive Stroke Certification. The other measures in the set include CSTK-02 Modified Rankin Score (mRS) at 90 Days; CSTK-03 Severity Measurement Performed for Subarachnoid Hemorrhage (SAH) and Intracerebral Hemorrhage (ICH) Patients (Overall Rate); CSTK-06 Nimodipine Treatment Administered. Although it is not required that these measures are reported in conjunction with each other, The Joint Commission develops measures in sets in order to provide as comprehensive a view of quality for a particular clinical topic as possible.
Developer Rationale: A neurological examination of all patients presenting to the hospital emergency department with warning signs and symptoms of stroke should be a top priority and performed in a timely fashion. Use of a standardized stroke scale or scoring tool ensures that the major components of the neurological examination are evaluated. Clinical practice guidelines from the American Heart Association/American Stroke Association recommend The National Institutes of Health Stroke Scale (NIHSS) as the preferred scoring tool for this purpose (Jauch, 2013). Scores obtained aid in the initial diagnosis of the patient, facilitate communication among healthcare professionals, and identify patient eligibility for various interventions and the potential for complications.

Numerator Statement: Ischemic stroke patients for whom an initial NIHSS score is performed prior to any acute recanalization therapy in patients undergoing recanalization therapy and documented in the medical record, OR documented within 12 hours of arrival at the hospital emergency department in patients who do not undergo recanalization therapy.
Denominator Statement: Ischemic stroke patients who arrive at this hospital emergency department (ED).
Denominator Exclusions:
- Patients less than 18 years of age
- Patients who have a Length of Stay greater than 120 days
- Patients with Comfort Measures Only documented on the day of or day after hospital arrival
- Patients admitted for Elective Carotid Intervention
- Patients who do not undergo recanalization therapy and are discharged within 12 hours of arrival at this hospital

Measure Type: Process
Data Source: Electronic Clinical Data, Paper Medical Records
Level of Analysis: Facility, Population : National

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☐ **Yes** ☒ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☒ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

- The developer provided a diagram of the relationship and summary statement: "Performance of NIHSS assessment for all ischemic stroke patients presenting to the hospital emergency department (ED) increases early detection and diagnosis of stroke and increases the identification of patients eligible for a treatment intervention, i.e., IV t-PA, IA t-PA, and helps predict the prognosis for outcome."
- The developer provided one guideline from the American Heart Association/American Stroke Association with two guideline statements for the emergency evaluation and diagnosis of acute ischemic stroke:
  - A complete evaluation/clinical assessment upon arrival to the ED is recommended for patients with suspected stroke. Evaluation/clinical assessment should include neurological examination. **Class I: Level of Evidence B**
  - The use of a stroke rating scale, preferably the NIH Stroke Scale, is recommended. "Use of a standardized assessment and stroke scale helps quantify the degree of neurological deficits, facilitate communication, identify the location of vessel occlusion, provide early prognosis, help select patients for various interventions, and identify the potential for complications." – based on three studies. **AHA/ASA Class I Recommendation: Level of Evidence B** (Data derived from a single randomized trial or nonrandomized studies).
- The developer presented two studies using NIHSS that were included in a the systematic review of 17 externally validated models to predict outcome following acute stroke. Both studies were consistent in concluding NIHSS ability to predict stroke outcomes.
- The developer also provided a recommendation from the American Heart Association/American Stroke Association's scientific statement for the use of telemedicine in the acute stroke setting when a stroke specialist is not immediately available. **Class I: Level of Evidence A**
  - The recommendation from the 2009 Review of the Evidence for the Use of Telemedicine within Stroke Systems of Care from the American Heart Association/American Stroke Association discusses conducting the NIH Stroke Scale using high-quality videoconferencing (HQ-VTC). The recommendation ("*The NIHSS-telestroke examination when administered by a stroke specialist using HQ-VTC, is recommended…*") is a Class I recommendation (signifying that it should be performed) with grade A evidence (meaning that it is based on data derived from multiple randomized clinical trials or meta-analyses).
- A systematic review regarding the reliability of the NIHSS is described. This evidence addresses the inter-rater reliability of NIHSS items. The summary of the evidence supports the reliability and feasibility of the NIHSS as a tool for rapidly assessing the effects of stroke. While the overall reliability of the NIHSS has been shown by several studies to be excellent, there is some disagreement as to which individual items have poor to fair reliability.
- The developer did not provide evidence to support the 12 hour timeframe for documenting that the NIH Score Scale was performed on patients who did not undergo recanalization therapy.

Exception to evidence: Not Applicable

Guidance from the Evidence Algorithm

Process measure (Box 3)→Systematic Review of the evidence within the AHA/ASA guideline development (Box 4)→QQC partially available (Box 4)→ Class I recommendation (Procedure/treatment SHOULD be performed/administered) implies that the evidence review concludes a moderate-high certainty that the net benefit is substantial (Box 5) → moderate (due to lack of detail on the quality and consistency of the evidence)

Questions for the Committee:

- *Does the evidence provided support the statement "Performance of NIHSS assessment for all ischemic stroke patients presenting to the hospital emergency department (ED) increases early detection and diagnosis of stroke and increases the identification of patients eligible for a treatment intervention, i.e., IV t-PA, IA t-PA, and helps predict the prognosis for outcome"?*
- *What is the relationship of documenting the NIH Stroke Scale on arrival to the ED or within 12 hours to patient outcomes?*
- *How strong is the evidence for these relationships?*
- *Are there, or could there be, performance measures of a related health outcome, OR evidence-based intermediate clinical outcomes, intervention/treatment?*
- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
- *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable without empirical evidence?*

Preliminary rating for evidence: ☐ High ☒ Moderate ☐ Low ☐ Insufficient

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer provided the following data for current performance:

	Pilot test (2012-2013)	1Q 2015	2Q 2015
# sites	66	38	50
# patients (den)	4,686	3,750	6,786
Hospital average rate	72%	84% (hospital mean rate)	85% (hospital mean rate)
Standard deviation	26%	15.3%	17.2%
Hospital median	70.8%	89%	89%
Range	0 - 100%	10^{th} percentile = 64% 25^{th} percentile = 80% 90^{th} percentile rate = 97%	10^{th} percentile = 73% 25^{th} percentile = 81% 90^{th} percentile rate = 96%
National aggregate rate		83%	83%

Disparities:

- The developer did not provide data on disparities from the measure as specified.
- The developer provided a summary of data from the literature that address disparities in stroke care.

Questions for the Committee:

- *Does the gap in care warrant a national performance measure?*
- *Does the literature on disparities provided by the developer address the focus of this measure which is determining*

whether there is documentation of the initial NIH Score Scale?

- *Are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **I am skeptical about the strength of the evidence presented. Why is the NIHSS recommended over the mNIHSS, which has fewer items and greater reliability? http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2729912/

**This process measure applies directly to the population of stroke patients. Use of this standardized stroke scale ensures that the appropriate assessment is completed. This tool enables effective communication of neurological status, prognosis, and treatment recommendation.

**Evidence supports NIHSS tool

**The evidence supporting the value of the NIH stroke scale and the value of performing this in the ED on suspected stroke patients is high. Class I data is provided along with a good systematic review. Several other joint commission process measures for stroke measure the percentage of patients receiving therapy in a timely fashion, which would require accurate and timely diagnosis, so there is a strong rationale for timely performance of the NIH stroke scale in the ED.

The major concern is the second inclusion group, which are patients who do not undergo recanalization therapy who have a documented NIH stroke scale within 12 hours. There does not appear to be a clear rationale provided for this nor strong support for this. It is not clear how performing an NIH stroke scale 11 hours after arrival would be of benefit, particularly since this time frame seems unrelated to last known well time. So if a patient arrived in the hospital 8 hours after onset of symptoms, and the NIHSS was performed 11 hours after that, this data would still be included even though the patient is 19 hours from last known well.

The timing of the implementation of the measure also appears vague even with the more acute patients and does not appear to be coordinated with other process and therapeutic measures which have strict time limits for implementation. Therefore, it is not clear how effective this measure will be in helping to drive critical therapeutic decision making.

1b. Performance Gap

Comments: **There does appear to be a gap in care and there is evidence of some ethnoracial disparity in receipt of the NIHSS (Bhattacharya 2013).

**Performance data was provided from a pilot test (66 sites n=4,686), 1Q2015 (38 sites, n=3,750), and 2Q2015 (50 sites, n=6,786) with a range of 72%-85% hospital mean rate of compliance. No data on subpopulations provided.

**Data demonstrates gap in care; no disparities data

**There is evidence overall of a small but meaningful performance gap based upon pilot data and published papers, representing a roughly 15% gap. There are also several publications provided to suggest that there remain racial disparities in application of the NIH stroke scale in a timely fashion.

1c. High Priority (previously referred to as High Impact)

Comments: **N/A

**N/A

**NA

**N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Medical record abstraction (paper or electronic). This is not an eMeasure.

Specifications:

- Nine data elements define the numerator. Patients are eligible for inclusion in the numerator if the NIH Score Scale is performed and documented prior to undergoing any acute recanalization therapy or documented within 12 hours of arrival at the emergency department for patients who do not undergo recanalization therapy.

- The data element "Initial NIHSS Score Performed" is defined as documentation of the first National Institutes of Health Stroke Scale (NIHSS) score that was done at this hospital. The NIHSS measures several aspects of brain function, including consciousness, vision, sensation, movement, speech, and language. The NIHSS serves several purposes, but its main use in clinical medicine is during the assessment of whether or not the degree of disability caused by a given stroke merits treatment with t-PA. Score documentation may range from 0 to 42.
- Allowable values for the NIHSS score for the measure numerator include: Yes or No/UTD.

- Eleven data elements define the denominator.
- Patients are excluded from the measure if they are under age 18, have a length of stay greater than 120 days, with comfort measures only documented on the day of or day after hospital arrival, are admitted for elective carotid intervention, or do not undergo recanalization therapy and are discharged within 12 hours of hospital arrival.
- ICD-10 codes provided for diagnoses and procedures. See attached spreadsheet Appendix A Table 8.1 (ischemic stroke).
- A detailed calculation algorithm is provided.
- Sampling monthly or quarterly is allowed; instructions for sampling are provided.
- A web-based data collection tool has been developed but is not described. Hospitals are not required to use this tool.
- The measures is specified at the hospital level of analysis.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method of reliability testing:

- Inter-rater reliability testing of 14 data elements (nine numerator data elements and four denominator data elements) was conducted at 12 sites with 281 total records from 2013. Percent agreement and Kappa scores were used to compare two sets of abstracted data. Analysis of inter-rater reliability is an appropriate test for data element reliability for abstracted data.
- As noted in the evidence section, above, the developers provided information regarding the reliability of the NIHSS itself (described as "excellent" based on a systematic review, though, there is some disagreement as to which individual items have poor to fair reliability).

Results of reliability testing:

- Percent agreement for the 14 data elements ranged from 71.5% (Discharge Time) to 99.3% (ED Patient). A Kappa score was calculated for Initial NIHSS Score Performed (kappa=0.95), ED Patient (kappa=0.96), Elective Carotid Intervention (kappa=-0.0043), and Warning Signs and Symptoms of Stroke (kappa=0.26). [*The data element Warning Signs and Symptoms of Stroke was removed based on feedback from the pilot sites because it was redundant with the principal diagnosis of ischemic stroke, consistently abstracted "Yes", and increased abstraction burden.*]
- The kappa statistic—appropriate for non-continuous variables—represents the proportion of agreement between two abstractors that is not explained by chance alone. Values for kappa range between -1.0 and 1.0. A value of 1.0 reflects perfect agreement; a value of 0 reflects agreement that is no better than what would be

expected by chance alone; a value less than zero reflects agreement that is worse than what would be expected by chance (and potentially, systematic disagreement). Percent agreement does not adjust for chance agreement and therefore should not be used alone to demonstrate reliability.

- Very low kappa values usually are interpreted as reflecting poor agreement. However, the kappa statistic is influenced by the prevalence of the attribute. The developers suggest that the low prevalence of patients with elective carotid intervention in the testing sample makes the value of kappa less informative and they instead rely on the percentage agreement value in their interpretation of this data element's reliability.

Guidance from the Reliability Algorithm:
Precise specifications (Box 1) →empirical testing with statistical tests as specified (Box 2) → reliability testing with patient-level data elements (Box 8) → assessed most critical data elements for numerator, less than half for denominator, and one exclusion, with the remaining denominator data elements obtained via administrative data (Box 9)→ high/moderate certainty the data elements are reliable (Box 10) → Moderate
Note: When only data element testing is performed (no measure score testing) the highest rating possible is moderate.

Questions for the Committee:

- *Does the low percent agreement for the time variables (arrival, initial NIHSS score, discharge) and the low kappa score for the exclusion, Elective Carotid Intervention, affect the reliability of the measure?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.
Specifications consistent with evidence in 1a. ☐ **Yes** ☒ **Somewhat** ☐ **No**
Specification not completely consistent with evidence: The developer did not provide evidence to support the 12 hour timeframe for documenting that the NIH Score Scale was performed on patients who did not undergo recanalization therapy.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING
Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☐ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- Empirical testing of the measure score:
  - Pilot testing data from October 1, 2012 – July 15, 2013 included information from 1,307 inpatient records from 66 hospitals. Additional testing data for the period Q1-Q2 2015 were obtained from 42

Joint Commission certified comprehensive stroke centers and included 65,389 inpatient records relevant to measures CSTK-1, CSTK-2, CSTK-3 and CSTK-6; only centers reporting at least 6 months of data and with ≥ 30 denominator cases were included in this dataset. It is unclear how many of the participating hospitals used electronic records versus paper records.

- Developers conducted several construct validation analyses, first hypothesizing a relationship between this measure and three other TJC stroke measures (specifically, that hospitals doing well on this measure also do well on the other measures) and then examining the degree of association between the measure results using the Pearson Correlation Coefficient. This is an appropriate method of score-level validation.
  - ***Hypothesis 1:*** Results from this measure should be positively correlated with results from the TJC CSTK-02 measure (Modified Rankin Score (mRS) at 90 Days). This hypothesis was tested using the pilot testing data and using the Q1-2 2015 data.
  - ***Hypothesis 2:*** Results from this measure should be positively correlated with results from the TJC CSTK-03 measure (Severity Measurement Performed for Subarachnoid Hemorrhage (SAH) and Intracerebral Hemorrhage (ICH) Patients (Overall Rate)). This hypothesis was tested using the Q1-2 2015 data.
  - ***Hypothesis 3:*** Results from this measure should be positively correlated with results from the TJC CSTK-06 measure (Nimodipine Treatment Administered). This hypothesis was tested using the Q1-2 2015 data.

- Face validity of data elements was reported but did not address face validity of the measure score as a representation of quality, as required for the criterion.
- Data element validity was assessed for accuracy and clarity by hospitals, but the data element validity criterion requires comparison to a gold standard.

Validity testing results:

- ***Hypothesis 1: Pilot testing data:***

Pearson Correlation Coefficients, N = 66 Prob > |r| under H0: Rho=0

	CSTK01	**CSTK02**
Correlation Coefficient	1.00000	0.25543
P value		0.0385

  - These results indicate a statistically significant positive correlation between the two measures, thus confirming the developers' hypothesis.
- ***Q1-2 2015 data: Hypotheses 1,2, and 3:*** Results for these analyses are presented in a correlation table. The correlation values are positive and statistically significant at the alpha=0.10 level, and therefore confirm the developers' hypotheses. These results found that this measure has the highest, significant correlation with CSTK 06: Nimodipine Treatment Administered.

Questions for the Committee:

- *Are the test samples adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

Patients are excluded from the measure if they are under age 18, have a length of stay greater than 120 days, with comfort measures only documented on the day of or day after hospital arrival, are admitted for elective carotid intervention, or do not undergo recanalization therapy and are discharged within 12 hours of hospital arrival.

The developers provide frequencies of exclusions for the pilot test data and the Q1-Q2 2015 data.

	Pilot test data	Q1-Q2 2015 data
<18	0.37%	--
LOS >120 days	0.11%	--
Admitted for Elective Carotid Intervention	0.34%	2.0%
Do not undergo re-canalization therapy and are discharged within 12 hours of arrival	0.14%	63.0%
Without warning signs and symptoms of stroke on arrival	5.81%	--
Comfort Measures Only	--	2.0%

-- Q1-Q2 2015 dataset already excluded patients <18 or LOS>120 days

Questions for the Committee:

- *In the Q1-Q2 2015 data sample more than half of the patients are excluded. Does the large number of exclusions for patients who do not undergo recanalization therapy and are discharged within 12 hours pose a threat to validity of the measure? Is this a reasonable exclusion for the measure?*
- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- See data described under Performance Gap, above.
- The difference between the 10th and 25th percentiles indicates some variation between providers.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*
- *Is there enough variation that audiences can identify high quality providers from lower quality providers?*

2b6. Comparability of data sources/methods:

- Only one data source (the medical record) is specified.

2b7. Missing Data

- The developer describes the handling of missing data in the measure specifications (S.22) but does not provide information on the frequency of missing data or potential impact on measure results.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → potential threats to validity mostly assessed (Box2) →empirical validity testing (Box 3) → measure score testing (Box 6) → sound conceptualization and hypotheses (Box 7) →high confidence that score are a valid indicator of quality(Box 8a) → High

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **The developer did not provide evidence to support the 12 hour timeframe.

**No information provided.
**Specifications somewhat consistent with evidence
**Validity is a major concern since less than 1% of patients were excluded for not undergoing recanalization therapy in the pilot sample, but in the Q1-Q2 2015 data, more than 60% of patients were excluded for this criteria alone. Elective carotid patients also were roughly 10 fold higher than the pilot data by percentage, so that 2/3 of patients overall were excluded, as opposed to less than 7% exclusions in the pilot data. This may speak to a larger than anticipated performance gap but raises real questions about validity.

2a2. Reliability Testing

Comments: **Did not address face validity of the measure score as a representation of quality and did not compare to a gold standard.
**Pilot testing data only had 66 participants. A minimum of 200 patients is required for validity testing against a gold standard. These samples may not be adequate to generalize for widespread implementation. The results do not demonstrate sufficient validity. However, I do agree that the score from this measure is an indicator of quality.
**Adequate scope and method

2b2. Validity Testing

Comments: **The exclusions are reasonable, but the percent discharged within 12 hours without undergoing re-canalization therapy seems very high (but those patients should definitely be excluded).
**These exclusions are not a threat to validity of this measure. For those who are discharged from the ED with minimal or no residual effect from stroke this measure is not relevant. The exclusions are consistent with the evidence. This does not inappropriately exclude patients groups from the measure. The exclusions do not outweigh the data collection burden.
**No threats to validity

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **I have concerns about the reliability of the measure, since the NIHSS itself has issues. The sample size of the population used for testing is not large.
**The low percent agreement for the time variables might be reflective of under reporting for discharge. This requires further investigation. The test sample is adequate to generalize for widespread implementation. The results demonstrate sufficient reliability so that differences in performance can be identified.
**Reliability testing appropriate

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is medical record abstraction: not all hospitals are able to generate the data electronically so a paper-based option is available
- Some data elements are in electronic form
- The measure is in use in the Joint Commission's Stroke Certification program.
- Data collection burden is significant:
  - Medical record review to abstract all data elements required for the measure set averaged 45 minutes per record.
  - Developer estimated " that the cost per case to abstract for this measure was approximately $3.50."

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Is the burden of data collection reasonable for a national performance measure?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
Comments: **The burden of data collection is very high for this measure, and while the cost per case ($3.50) may seem low, the total would likely be several million dollars.
**The required data elements may not be routinely generated and used during care delivery, however, they are built in the electronic health record flow sheets in most institutions. The burden of data collection is reasonable for a national performance measure. The data collection strategy is ready to be put into operational use. The large EHR vendors have this built for immediate implementation.
**Data collection burden high but reasonable
**Feasibility is moderate. Should be readily implemented and there are paper options for centers that do not have an EHR. However, the developer estimates roughly 45 minutes to abstract all data elements from the medical record review, which seems high if the measure is appropriately implemented and flagged at centers which wish to be stroke centers. However, if true, the developer estimate of the added cost for abstracting this measure at $3.50, which seems quite low since this would be prorated at roughly $4.67 per hour, which is nearly $3 below the federally mandated minimum wage for a single worker, without accounting for any non-personnel costs such as time or opportunity costs.
The developer also notes improvement in implementation from pilot test of 72% to Q1 national average of 83%, but this was also coincident with an exclusion rate of less than 7% in the pilot data and 67% in the Q1-Q2 national data, so it is not clear how well this reflected feasibility if the majority of patients were excluded to achieve this increase in utilization.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:
- The Joint Commission Disease-Specific Care Certification for Comprehensive Stroke Centers is a certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
- Additional planned uses include public reporting and external benchmarking to other organizations, although no timeframe was noted.

Improvement results:
- The developers note improvement from the pilot test (average results 72%) to a national rate of 83% for 2Q 2015.

Potential harms:
- There were no unintended negative consequences reported or detected during testing or since implementation of the measure specifications.

Feedback :
- New measure with relatively little use; 50 hospitals nationwide implemented this measure as of 2Q 2015.

Questions for the Committee:
- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

- *Are hospitals likely to improve results quickly to overall high performance?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences

Comments: **Not sure that the benefits outweigh the costs, given the flawed scale, high burden of data collection, and lack of evidence for the measure as specified.
**The performance results can be used to further the goal of high-quality, efficient healthcare by requiring the public reporting of these data. Currently it is only required as part of JC certification. Hospitals are likely to improve results quickly to overall performance if this indicator is required. The benefits of the measure outweigh the burden of data collection.
**Not publicly reported but Joint Commission accountability; benefits outweigh unintended consequences
**Usability remains a bit of a question as implementation has been somewhat slow, but this likely relates to the overall stringent requirements for becoming a joint commission stroke center and is not likely specific to usability of this measure. The developer again note improvement in utilization from 72% in the pilot data to 83% in the Q2 national data, and I again note that this was coincident with a dramatic increase in exclusions which represented the majority of patients abstracted.

Criterion 5: Related and Competing Measures

Related or competing measures

- NQF #2396 : Carotid Artery Stenting: Evaluation of Vital Status and NIH Stroke Scale at Follow Up

Harmonization:

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Title: CSTK-01: National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure title

Date of Submission: 1/15/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to

demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

Subcriterion 1a. Evidence to Support the Measure Focus

The measure focus is a health outcome or is evidence-based, demonstrated as follows:

- Health outcome:[3] a rationale supports the relationship of the health outcome to processes or structures of care.
- Intermediate clinical outcome, Process,[4] or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence[5] that the measure focus leads to a desired health outcome.
- Patient experience with care: evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information OR that patient experience with care is correlated with desired outcomes.
- Efficiency:[6] evidence for the quality component as noted above.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement.

5. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

6. Measures of efficiency combine the concepts of resource use and quality (NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1. This is a measure of:

Outcome

☐ Health outcome: Click here to name the health outcome

Health outcome includes patient-reported outcomes (PRO, i.e., HRQoL/functional status, symptom/burden, experience with care, health-related behaviors)

☐ Intermediate clinical outcome: Click here to name the intermediate outcome

☐ Process: NIHSS score performed and documented for ischemic stroke patients prior to intravenous thrombolytic (IV t-PA) therapy, intra-arterial thrombolytic (IA t-PA) therapy, or mechanical endovascular reperfusion therapy, or performed within 12 hours of hospital arrival for ischemic stroke patients who do not undergo these therapies.

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME PERFORMANCE MEASURE *If not a health outcome, skip to 1a.3*

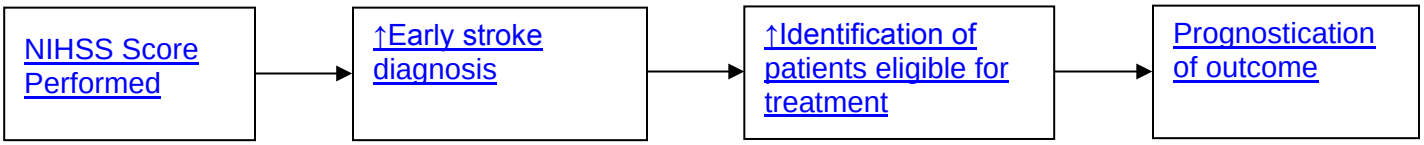
1a.2. Briefly state or diagram the linkage between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

Note: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the linkages between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



Performance of NIHSS assessment for all ischemic stroke patients presenting to the hospital emergency department (ED) increases early detection and diagnosis of stroke and increases the identification of patients eligible for a treatment intervention, i.e., IV t-PA, IA t-PA, and helps predict the prognosis for outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***
☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***
☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***
☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):
Jauch, E. C., J. L. Saver, H. P. Adams, Jr., A. Bruno, J. J. Connors, B. M. Demaerschalk, P. Khatri*, et al.* Guidelines for the Early Management of Patients with Acute Ischemic Stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association. *Stroke.* 2013;44:13.

URL:
http://stroke.ahajournals.org/content/early/2013/01/31/STR.0b013e318284056a.full.pdf+html

Schwamm LH, Holloway RG, Amarenco P, Audebert HJ, Bakas T, Chumbler NR, Handschu R, Jauch EC, Knight WA IV, Levine SR, Mayberg M, Meyer BC, Meyers PM, Skalabrin E, Wechsler LR. A review of the evidence for the use of telemedicine within stroke systems of care: A Scientific Statement from the American Heart Association/American Stroke Association. *Stroke.* 2009;40:2616-2634.

URL:
http://stroke.ahajournals.org/content/40/7/2635.full.pdf+html

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Jauch, 2013. Emergency Evaluation and Diagnosis of Acute Ischemic Stroke: Neurological Examination and Stroke Scale/Scores (Page 13)

Class I

1. An organized protocol for the emergency evaluation of patients with suspected stroke is recommended (*Level of Evidence: B*). The goal is to complete an evaluation to begin fibrinolytic treatment within 60 minutes of the patient's arrival in an ED. Designation of an acute stroke team that includes physicians, nurses, and laboratory/radiology personnel is encouraged. Patients with stroke should have a careful clinical assessment, including neurological examination.
2. The use of a stroke rating scale, preferably the NIHSS, is recommended (*Level of Evidence: B*).

Schwamm, 2009. Acute Stroke Evaluation, Including the Hyperacute and Emergency Department Phases: Acute Stroke Setting, Including Thrombolytic Evaluation (Page 2623)
Class I

1. The NIHSS-telestroke examination when administered by a stroke specialist using HQ-VTC, is recommended when an NIHSS-bedside assessment by a stroke specialist is not immediately available for patients in the acute stroke setting, and this assessment is comparable to an NIHSS-bedside assessment (*Level of Evidence: A*).

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Class I: Procedure/treatment SHOULD be performed/administered.

Level A: Data derived from multiple randomized clinical trials or meta-analyses.
Level of Evidence B: Data derived from a single randomized trial or nonrandomized studies.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

ACCF/AHA Classification of Recommendation and Level of Evidence
Classification Types
Class II a: It is reasonable to perform procedure/administer treatment.
Class II b: Procedure/Treatment may be considered.
Class III: Procedure/Treatment not helpful/no proven benefit/may be harmful.

Level of Evidence
Level C: Only consensus opinion of experts, case studies, or standard of care.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**
Same as 1a.4.1

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?
☐ Yes → ***complete section 1a.7*** Schwamm, 2009
☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION
1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE
1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Teale EA, Forster A, Munyombwe T, Young JB. A systematic review of case-mix adjustment models for stroke. *Clinical Rehabilitation.* 2102;26(9):771-786.

URL:
http://web.b.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=1e41cde3-0601-47ea-b3b7-e5c8e21b8bd2%40sessionmgr113&vid=2&hid=122

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Same as 1a.6.1

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE
1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?
The information in the following questions in this section is based upon the studies included in reference to the recommendation noted above from the 2009 AHA/ASA Scientific Statement, "A Review of the Evidence for the Use of Telemedicine Within Stroke Systems of Care" and also the systematic review conducted by Teale, et. al., in 2012.

Schwamm, 2009
An evidence table supporting the reliability of NIHSS studies as published in the AHA/ASA Scientific Statement is summarized below. An analysis of 9 studies demonstrated that the percentage of NIHSS items with excellent inter-rater reliability ranges from 31% to 38%.

Goldstein, 1989

Kappa	> 0.60 = Excellent
% Excellent	5/13 (38%)
% Moderate	4/13 (31%)
% Poor	4/13 (31%)

Brott, 1989

Kappa	> 0.80 = Excellent
% Excellent	4/13 (31%)
% Moderate	9/13 (69%)
% Poor	0/13 0%

Shafquat, 1999

Kappa	> 0.75 = Excellent
% Excellent	4/13 (31)
% Moderate	7/13 (54%)
% Poor	2/13 (15%)

Meyer, 2005

Kappa	> 0.75 = Excellent
% Excellent	10/15 (67%)

Meyer, 2005

Kappa	> 0.75 = Excellent
% Excellent	10/15 (67%)

% Moderate	3/15 (20%)	**% Moderate**	5/15 (33%)
% Poor	2/15 (13%)	**% Poor**	0/15 (0%)

Handschu, 2003		*Handschu,2003*	
Kappa	Weighted	**Kappa**	Weighted
% Excellent	13/13 (100%)	**% Excellent**	12/13 (92%)
% Moderate	0/13 (0%)	**% Moderate**	1/13 (8%)
% Poor	0/13 (0%)	**% Poor**	0/13 (0%)

LaMonte, 2004		LaMonte, 2004	
Kappa	r > 0.5 = Good	**Kappa**	r > 0.5 = Good
% Excellent	6/15 (40%)	**% Excellent**	7/15 (47%)
% Moderate	7/15 (47%)	**% Moderate**	5/15 (33%)
% Poor	2/15 (13%)	**% Poor**	3/15 (20%)

Teale, 2012

Two studies using NIHSS were included in the systematic review of 17 externally validated models to predict outcome following acute stroke:

Study	**Sample Size**	**Outcome Assessed**	**Model Performance**
Muir, et. al., 1996	408	Alive at home vs. dead or In care at 3 months	Prediction of poor outcome: Sensitivity 71%, Specificity 90%
Lai, et. al., 1998	184	Bartel Index (treated as Interval data) at 1, 3, 6 months	R^2 = 0.56 at 1 month <0.5 at 3 and 6 months

This systematic review concluded that many existing prognostic models in stroke predict mortality or dependency; however, these endpoints may be of less interest to individual patients and their caregivers than more complex rehabilitation outcomes. Alternative modeling approaches for case-mix adjustment that focus on the recovery of the individual post-stroke may improve existing models. Further research is needed to explore alternative modeling.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

An overall grade for the quality of quoted evidence was not assigned for either review.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

An overall grade for the quality of quoted evidence was not assigned for either review.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range
1989 – 2005 (Schwamm)
1996 – 1998 (Teale)

QUANTITY AND QUALITY OF BODY OF EVIDENCE
1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)
Schwamm, 2009 - 6 observational studies
Teale, 2012 – 2 observational studies

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Overall quality of evidence across studies appears to be high: Eight studies are noted here:

Study	Quality of Study
Goldstein, 1989	• Retrospective observational study • n = 20 stroke patients • Design: The inter-observer reliability of the NIHSS employed in several multicenter stroke trials was investigated. Twenty stroke patients were rated with this scale by four clinical stroke fellows. Each patient was independently evaluated by one pair of observers. • Confidence of effect: The degree of inter-rater agreement for each item on the NIHSS was determined by calculation of the *k* statistic. Inter-observer agreement was moderate to substantial for 9 of 13 items. The NIHSS compares favorably with other scales for which such comparisons can be made. • Potential limitation: Small 'n'.
Brott, 1989	• Prospective observational study • n = 24 stroke patients; n = 65 stroke patients • Design: A 15-item neurological scale was evaluated for test-retest reliability and inter-rater reliability (n=24). Second the scale was field-tested as part of a prospective therapy (naloxone) study of 65 patients with acute cerebral infarction. Finally, validity was assessed by correlating the scale scores with two other measures of stroke severity: computed tomography (CT) measurement of volume of infarction and clinical outcome at 3 months. • Confidence of effect: Inter-rater reliability for the scale was found to be high (mean *k*=0.69) and test-retest reliability was also high (mean *k*=0.66-0.77). Test-retest reliability did not differ significantly among a neurologist, a neurology house officer, a neurology nurse, or an emergency department nurse. The stroke scale validity was assessed by comparing the scale scores to the patient's infarct size as measured by CT at 1 week and to the patient's clinical outcome at 3 months. These correlations (scale-lesion size *r*=0.68, scale-outcome *r*=0.79) suggested acceptable examination and scale validity. • Potential limitation: Small 'n'. Also, reliability results were probably improved by having one examiner and three observers together scoring each patient on two occasions (albeit scoring independently and without knowledge of the other scorer's opinions) rather than having each examination team member by himself examining each patient on two occasions. This technique means that changes in patient performance induced by fatigue or practice effect are less likely to cause differences between the examinations. Reliability might have been improved in

	that the examiners' scores for the second examination may have been influenced by recollections of the first examination (though examiners were blinded to the first examination scores). Reliability may also be anticipated to be lower for other examiners, as the examination team members of this study developed the scale and so became experienced together before the reliability testing.
Muir, 1996 (Teale, 2012)	• Prospective observational study • n = 408 patients studied; 373 stroke patients • Design: A single observer scored consecutive admissions to an acute stroke unit on the NIHSS, the Canadian Neurological Scale, and the Middle Cerebral Artery Neurological Score. Guy's prognostic score was determined from clinical data. Outcome at 2, 3, 6, and 12 months was categorized as good (alive at home) or poor (alive in care or dead). Predictive accuracy of the variables was compared by receiver operating characteristic curves and stepwise logistic regression. • Confidence of effect: The three stroke rating scales each predicted 3-month outcome with an accuracy of .79 or greater. The NIHSS provided the most prognostic information: sensitivity to poor outcome, .71 (95% confidence interval [CI], .64 to.79); specificity, .90 (95% CI, .86 to .94); and overall accuracy, .83 (95% CI, .79 to .87). Logistic regression showed that the NIHSS added significantly to the predictive value of other scores. No score added useful information to the NIHSS. A cut point of 13 on the IHSS best predicted 3-month outcome. • Potential limitation: The availability of staff to perform NIHSS may limit use to specialist settings.
Lai, 1998 (Teale, 2012)	• Prospective observational study • n = 184 stroke patients recruited for the Kansas City Stroke Study. • Design: This study compared the ability of two stroke impairment scales, Orpington Prognostic Scale and National Institutes of Health (NIH) Stroke Scale, to predict disability as measured by the Barthel activities of daily living (ADL) Index and higher level of self-reported physical functioning as measured by the SF-36 physical functioning index (PFI) at 1, 3, and 6 months after stroke. All patients were prospectively evaluated using standardized assessments at enrollment (within 14 days of stroke onset) and followed at 1, 3, and 6 months after stroke. Coefficient of determination (*R*2) was used to assess the ability of the 2 stroke scales to prognosticate outcomes. • Confidence of effect: Means and SDs of the Orpington Prognostic Scale and NIH Stroke Scale measured at baseline were 3.661.31and 5.564.58, respectively. The Spearman's rank correlation between the 2 baseline measures was 0.83 (*P*50.0001). The Orpington Prognostic Scale and the NIH Stroke Scale explained well the variance in Barthel ADL Index (*P*,0.001). The amount of variance in Barthel ADL Index and SF-36 PFI, which were

	explained by both stroke severity measures, decreased over time. • Potential limitation: Sample included mostly mild and moderate strokes.
Shafqat, 1999	• Prospective observational study • n = 20 ischemic stroke patients • Design: One bedside and 1 remote NIHSS score were independently obtained on each of 20 patients with ischemic stroke. The bedside examination was performed by a stroke neurologist at the patient's bedside. The remote examination was performed by a second stroke neurologist through an interactive high-speed audio-video link, assisted by a nurse at the patient's bedside. Kappa coefficients were calculated for concordance between bedside and remote scores. • Confidence of effect: Remote assessments took slightly longer than bedside assessments (mean 9.70 versus 6.55 minutes, *P*,0.001). NIHSS scores ranged from 1 through 24. Based on weighted k coefficients, 4 items (orientation, motor arm, motor leg, and neglect) displayed excellent agreement, 6 items (language, dysarthria, sensation, visual fields, facial palsy, and gaze) displayed good agreement, and 2 items (commands and ataxia) displayed poor agreement. Total NIHSS scores obtained by bedside and remote methods were strongly correlated (*r*50.97, *P*,0.001). • Potential limitation: Small 'n'. Also, no consistent pattern of inter-rater agreement was noted in the study which may be due to natural variability.
Meyer, 2005	• Prospective observational study • n = 25 patients with stroke symptoms • Design: Patients were examined both at bedside and via telemedicine by 2 NIHSS-certified neurologists. One examiner (bedside) examined the patients at bedside; the second examiner (remote) performed scale evaluations via the STRokE Doc (Stroke Team Remote Evaluation using a Digital Camera) system. The remote neurologist directed the examination assisted by the onsite neurologist rather than by a nonphysician assistant. • Confidence of effect: Feasibility was shown with all NIHSS-telestroke examinations (25 of 25, 100%) performed successfully with wireless telemedicine. NIHSS-bedside and NIHSS-telestroke scores ranged from 1 to 16. Intra-class correlation coefficient was 0.94 for NIHSS and 0.95 for modified NIHSS using weighted *K* coefficients, this trial showed the 67% of NIHSS items and 82% of modified NIHSS items had excellent agreement. • Potential limitation: Small 'n'.
Handschu,2003	• Prospective observational study • n = 41 stroke patients • Design: Acute stroke patients admitted to a stroke unit were examined on admission in the emergency room. Standardized examination was performed by use of the NIHSS (German version) via telemedicine and compared

	with bedside application. NIHSS-bedside and NIHSS-telestroke scores were performed by stroke neurologists and assisted by a trained medical student for the remote evaluations and ranged from 1 to 24. • Confidence of effect: Weighted *K* results showed excellent reliability for all 13 items in 41 patients examined within 36 hours of stroke onset (weighted *K*=1.0) and in 12 patients examined within 6 hours (weighted *K*=0.92). These 2 reports extended the feasibility and reliability of NIHSS-telestroke administered by telemedicine to the acute hospital setting and time period when therapeutic decisions are generally made. • Potential limitation: Minor issues with video (n=2), audio (n=5), and lighting (n=3) required repetition of the NIHSS-telestroke score in 2 cases.
LaMonte, 2004	• Prospective observational study • n = unknown • Design: Validity and reliability were tested by comparing neurologic examination scores obtained using our wireless system, which transmits video of a patient from a moving ambulance to desktop computers, with those obtained using the National Institute of Neurological Disorders and Stroke training videotape. TeleBAT validity and good interrater reliability were defined a priori as a kappa statistic of r > 0.5. We compared the average time to treatment for our TeleBAT-evaluated intervention group with that for our control group. The intervention group consisted of two actor patients with stroke mimicking 12 stroke scenarios and evaluated using TeleBAT. The control group consisted of patients with stroke evaluated and treated with rt-PA on arrival to the emergency department. Data were analyzed using standard t test. • Confidence of effect: National Institutes of Health Stroke Scale items calculated by the neurologists suggest TeleBAT is valid for assessing patients with stroke remotely. Inter-rater reliability was high: the neurologists gleaned the same information from TeleBAT transmissions. Kappa values for both validity and reliability exceeded 0.5. The mean time to treatment for patients assessed by TeleBAT was 17 +/- 4 minutes compared with 33 +/- 17 minutes for the control group (P = .0033). • Potential limitation: Future studies should use a randomized design with patients with acute stroke.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

The evidence supports the reliability and feasibility of the NIHSS as a tool for rapidly assessing the effects of stroke. Minimal amounts of training are required to reliably administer this tool, even by non-neurologists, which takes less than 10 minutes to complete. In comparison with other tools and an objective measurement of infarction volume the tool has been shown to record a highly valid assessment of stroke severity (r=0.68) and later effects (r=0.79) (Brott, 1989).

While the overall reliability of the NIHSS has been shown by several studies to be excellent, there is some disagreement as to which individual items have poor to fair reliability. The original authors evaluated the scale reliability using kappa statistic and found that, while most items had good to excellent reliability (Cronbach alpha > 0.5), two items, dysarthria and consciousness, rated fair to poor (Brott, 1989). In another analysis (Goldstein 1989), it was found that 13 of the 15 items making up the NIHSS showed no statistical difference between the observers. The observers had poor agreement on score in determining facial palsy and limb ataxia (alpha< 0.3).

Study	Results
Goldstein, 1989	Reliability was found to be excellent overall and moderate to excellent for most individual scales.
Brott, 1989	Reliability was found to be excellent overall and moderate to excellent for most individual scales.
Muir, 1996 (Teale, 2012)	Baseline NIHSS best predicts 3-month outcome.
Lai, 1998 (Teale, 2012)	Orpington Prognostic Scale has a slightly higher predictive value for mild to moderate strokes compared with that of the NIHSS.
Shafqat, 1999	NIHSS is also reliable when used to assess patients through a remote television link.
Meyer, 2005	Overall reliability was found to be excellent.
Handschu,2003	Excellent reliability for all 13 items.
LaMonte, 2004	TeleBAT (ambulance transport system) is valid and reliable means of evaluating stroke neurologic deficits (i.e., NIHSS).

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?
The evidence indicates that the NIHSS is a reliable, quick, and cost-effective tool for the assessment of stroke severity when performed at the bedside or remotely via telemedicine. Because “time is brain”, the use of the NIHSS performed in a timely and uniform fashion promotes early detection and treatment of stroke. However, the evaluation of stroke severity depends upon the ability of the observer to accurately and consistently assess the patient. Inter-observer variability is an inherent harm that can negatively impact score reliability. Reliability of observer assessment has been shown to increase after viewing instructional video tape recordings made by the authors.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
CSTK-01_Measure__Evidence_6.5.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
A neurological examination of all patients presenting to the hospital emergency department with warning signs and symptoms of stroke should be a top priority and performed in a timely fashion. Use of a standardized stroke scale or scoring tool ensures that the major components of the neurological examination are evaluated. Clinical practice guidelines from the American Heart Association/American Stroke Association recommend The National Institutes of Health Stroke Scale (NIHSS) as the preferred scoring tool for this purpose (Jauch, 2013). Scores obtained aid in the initial diagnosis of the patient, facilitate communication among healthcare professionals, and identify patient eligibility for various interventions and the potential for complications.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Pilot Test Findings:
During the six-month pilot test (October 1, 2012 – March 31, 2013), sixty-six sites submitted data for 10,218 completed patient records. For this measure, 4686 cases were assigned an ICD-9-CM Principal Diagnosis Code for ischemic stroke at discharge, and 3356 of these cases were captured in the numerator population. The performance rates varied widely across sites for this measure with results ranging from a low of 0% to a high of 100%. The average rate for all sites collecting data for this measure was 72%, indicating a potential performance gap of approximately 28% if the optimal rate is 100%.

In January, 2015, The Joint Commission implemented data collection for the comprehensive stroke (CSTK) measure set to meet performance measurement requirements for its Comprehensive Stroke Certification Program. Below is the specified level of analysis for CSTK-01 NIHSS Score Performed for Ischemic Stroke Patients for the two quarters of data received to date for this measure.

1Q 2015: 3750 denominator cases; 4536 numerator cases; 38 hospitals; 0.82762 national aggregate rate; 0.84306 mean of hospital rates; 0.15344 standard deviation; 0.96591 90th percentile rate; 0.94709 75th percentile rate/upper quartile; 0.88883 50th percentile rate/median rate; 0.79592 25th percentile rate/lower quartile; and, 0.64103 10th percentile rate.

2Q 2015: 6786 denominator cases; 5740 numerator cases; 50 hospitals; 0.83112 national aggregate rate; 0.85148 mean of hospital rates; 0.17241 standard deviation; 0.95628 90th percentile rate; 0.93636 75th percentile rate/upper quartile; 0.88877 50th percentile rate/median rate; 0.81231 25th percentile rate/lower quartile; and, 0.73039 10th percentile rate.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
Citation:
Fonarow, GC, Reeves MJ, Smith EE, Saver JL, Zhao X, Olson DW, Hernandez AF, Peterson ED, Schwamm LH. Characteristics, performance measures, and in-hospital outcomes of the first one million stroke and transient ischemic attack admissions in Get With The Guidelines-Stroke.

Summary:
Data analyzed: GWTG-Stroke Program 2003 to 2009; 1392 U.S. hospitals; 1,000,000 patients. Study limitations: It was not possible to account for stroke severity in these analyses because the NIHSS is inconsistently documented in the database, and so NIHSS

inclusion in the multivariable models may have introduced significant selection bias. Circ Cardiovasc Qual Outcomes. 2010;3:291-302.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This is the initial submission of this measure.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
Karve SJ, Balkrishnan R, Mohammad YM, Levine DA. Racial/ethnic disparities in emergency department waiting time for stroke patients in the United States. J Stroke Cerebrovasc Dis. 2011 Jan-Feb;20(1):30-40.

Data analyzed: 1997-2000 and 2003-2005; national sample; 543 ischemic and hemorrhagic stroke patients. Race was significantly associated with emergency department waiting time (EDWT) > 10 min (whites, 55% [referent]; blacks, 70% [P=.03}; Hispanics, 62% [P=.53]. These differences persisted after adjustment (blacks: odds ratio [OR]=2.08, 95% confidence interval [CI] = 1.05-4.09; Hispanics: OR=1.07, 95% CI=0.52-2.22. Blacks, but not Hispanics, had significantly longer EDWT than whites. The longer EDWT in black stroke patients may lead to treatment delays and suboptimal stroke care.

_____

Bhattacharya P, Mada F, Salowich-Palm L, Hinton S, Millis S, Watson SR, Chaturvedi S, Rajamani K. Department of neurology and Stroke Program, Wayne State University School of Medicine, Detroit, Michigan. Are racial disparities in stroke care still prevalent in certified stroke centers? J Stroke Cerebrovasc Dis. 2013. 22(4):383-388.

A retrospective chart review of 574 patients (25.1% African American) with ischemic stroke admitted to five Joint Commission certified primary stroke centers and five non-certified hospitals was conducted. Whites were more likely to arrive by emergency transport services (65.5% vs.51.1%; P= 0.004) to be evaluated by a stroke team (19.1% vs. 7.7%; P=0.001), and to have documented National Institutes of Health Stroke Scale (NIHSS) score (40.2% vs. 29.9%; P=0.03); however, the number of white and black patients who received IV t-PA was not statistically different (2.1% in African Americans, 3.5% in Caucasians; P=0.40).

_____

Boehme AK, Siegler JE, Mullen KT, Albright KC, Lyerly MJ, Monlezun DJ. Jones EM, Tanner R, Gonzales NR, Beasley TM, Grotta JC, Savitz SI, Martin-Schild S. Racial and gender differences in stroke severity, outcomes. and treatment in patients with acute ischemic stroke. J Stroke Cerebrovasc Dis. 2014. 23;4:e255-e261.

Data analyzed: between 2004 and 2011; 2 academic medical centers in the United states; 4925 patients. NIHSS on admission, median (range): 7 (0-40), black men (n=953) and 6 (0-40), white men (n=1626), P=.2748; 7 (0-40), black women (n=989) and 8 (0-40), white women (n=1357), P=.0144; 4-Group comparison P value <.0001. Median NIHSS score on admission was significantly different across all 4 groups due to a significant difference in admission NIHSS score between black women and white women. The difference in admission NIHSS score was not significant between black men and white men. After adjusting for age and glucose on admission in the analysis of variance model, the difference in admission NIHSS score between white women and black women was no longer significant (P=.1673).

_____

Newman-Toker DE, Moy E, Valente E, Coffey R, Hines AL. Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. Diagnosis. 2014; 1(2):155-166.

Cross-sectional analysis using linked inpatient discharge and ED visit records from the 2009 Healthcare Cost and Utilization Project State Inpatient Databases and 2008-2009 State ED Databases across nine U.S. states. The overall prevalence of misdiagnosis followed by delayed hospital admission after ED discharge ranges between 1.2% (probable missed strokes) and 12.7% (potential missed strokes) of all stroke admissions. Missed ischemic strokes (n=1435) and transient ischemic attack (n=402) were linked to nonspecific presenting symptoms of headache or dizziness. Bedside methods for identifying high-risk patients offer evidence-based, cost-effective strategies to reduce misdiagnosis. Odds of probable misdiagnosis were lower among men (OR 0.75), older individuals (18-44 years [base]; 45-64:OR 0.43; 65-74:OR 0.28; > 75:OR 0.19), and Medicare (OR 0.66) or Medicaid (OR 0.70) recipients

compared to privately insured patients. Odds were higher among Blacks (OR 1.18), Asian/Pacific Islanders (OR 1.29), and Hispanics (OR 1.30). Odds were higher in non-teaching hospitals (OR 1.45) and hospitals with low-volume ED visits.

_____

Leira EC, Hess DC, Torner JC, Adams HP Jr. Rural-urban differences in acute stroke management practices. Arch Neurol. 2008;65(7):887-891.

A PubMed search was conducted to identify all articles from 1997-2007 that addressed acute stroke, paramedics, ambulances, emergency services, and interhospital transportation pertaining to U.S. rural, urban, or nonurban environment. Expertise and training of rural paramedics and lack of consultant neurologists were two identified problems. The researchers concluded that acute stroke management practices in rural areas are sub-optimal, which creates an unacceptable health disparity between urban stroke patients and their rural counterparts, who constitute 25% of the U.S. population. A "hub-and-spoke" system, in which a large comprehensive system provides multimodal assistance to a group of rural hospitals, is an efficient way to organize and improve support.

_____

Gebhardt JG, Norris TE. Acute stroke care at rural hospitals in Idaho: challenges in expediting stroke care. The Journal of Rural Health. 2006;22(1):88-91.

Using a standardized questionnaire, a telephone survey of hospital staff at 21 rural hospitals in Idaho was performed. The survey focused on acute stroke care practices and strategies to expedite stroke care. The median number of stroke patients each year was 23.3. Approximately 67% of hospitals had implemented a clinical pathway for stroke and 80.0% had provided staff with stroke-specific training. No hospitals surveyed had a designated stroke team.

_____

Okon NJ, Rodriguez DV, Dietrich DW, Oser CS, Blades LL, Burnett AM, Russell JA, Allen MJ, Chasson L, Helgerson SD, Gohdes D, Harwell TS. Availability of diagnostic and treatment services for acute stroke in frontier counties in Montana and Northern Wyoming. The Journal of Rural Health. 2006;22(3):237-241.

In 2004, hospital medical directors or their designees were mailed a survey about the availability of diagnostic technology, programs, and personnel for acute stroke care. Fifty-eight of 67 (87%) hospitals responded to the survey; 79% were located in frontier counties with an average bed size of 18 (11SD). Of the hospitals in frontier counties, 44% reported emergency medical services pre-hospital stroke identification programs, 39% had 24-hour computed tomography capability, 44% had an emergency department stroke protocol, and 61% had a recombinant tissue plasminogen activator protocol.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. Demonstrated high priority aspect of healthcare

Affects large numbers, A leading cause of morbidity/mortality, Severity of illness

1c.2. If Other:

1c.3. Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.

Each year, ~ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. Approximately 55,000 more women than men have a stroke. Of all strokes, 87% are ischemic strokes.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. According to 2010 mortality data, one of every 20 deaths in the United States is attributable to stroke; 60% of stroke deaths are women.

Stroke is also a leading cause of serious long-term disability in the United States (Survey of Income and Program Participation, a survey of the US Census Bureau). In 2010, stroke was among the top 18 diseases contributing to years lived with disability; of these 18 causes, only the age-adjusted rates for stroke increased significantly between 1990 and 2010 (P<0.05). Data from the National Heart, Lung and Blood Institute (NHLBI) revealed that 50% of ischemic stroke survivors age > 65 years had some hemiparesis; 46% had cognitive deficits, 35% experienced depressive symptoms; 30% were unable to ambulate without assistance; 26% were dependent in activities of daily living; 19% had aphasia; and, 26% were institutionalized in a nursing home. Among Medicare patients discharged from the hospital after stroke, approximately 45% are discharged to home (32% returning to home use home healthcare services), 24% to an inpatient rehabilitation facility, and 31% discharged to a skilled nursing facility. The mean lifetime cost of ischemic stroke, including inpatient care, rehabilitation, and follow-up as necessary for residual deficits are estimated at $140,048 per person.

Strokes are the most common cause of new focal neurological deficits; therefore it is essential that patients presenting to the hospital emergency department with warning signs and symptoms of stroke receive a brief but thorough initial neurological examination. The use of a standardized neurological examination ensures that the major components of a neurological examination are performed in a timely and uniform manner. The National Institutes of Health Stroke Scale (NIHSS) Score is the preferred scoring tool in the United States for this purpose. The NIHSS may be performed rapidly, has demonstrated utility, and may be administered by a broad spectrum of healthcare providers.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Jauch, E. C., J. L. Saver, H. P. Adams, Jr., A. Bruno, J. J. Connors, B. M. Demaerschalk, P. Khatri, et al. Guidelines for the Early Management of Patients with Acute Ischemic Stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association. Stroke. 2013;44:10-11.

Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

Not applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*

Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

Care Coordination, Functional Status, Health and Functional Status : Functional Status

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

http://www.jointcommission.org/assets/1/6/CSTKManual2015August.pdf

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Copy_of_AppendixACSTKTables_ICD10codes-635878789321771970.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not Applicable

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Ischemic stroke patients for whom an initial NIHSS score is performed prior to any acute recanalization therapy in patients undergoing recanalization therapy and documented in the medical record, OR documented within 12 hours of arrival at the hospital emergency department in patients who do not undergo recanalization therapy.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Nine data elements are used to calculate the numerator. Data elements and definitions:
• Arrival Date - The earliest documented month, day, and year, the patient arrived at the hospital.

• Arrival Time - The earliest documented time (military time) the patient arrived at the hospital.

• ICD-10-PCS Other Procedure Code Date – The month, date, and year when the other procedure(s) was performed.

• ICD-10-PCS Other Procedure Code Time - The time (military time) when the other procedure(s) was performed.

• ICD-10-PCS Principal Procedure Code Date – The month, date, and year when the principal procedure was performed.

• ICD-10-PCS Principal Procedure Code Time - The time (military time) when the principal procedure was performed.

• Initial NIHSS Score Date – The month, date, and year the NIHSS score was first performed at the hospital.

• Initial NIHSS Score Performed – Documentation of the first National Institutes of Health Stroke Scale (NIHSS) score that was done at this hospital. The NIHSS measures several aspects of brain function, including consciousness, vision, sensation, movement, speech, and language. The NIHSS serves several purposes, but its main use in clinical medicine is during the assessment of whether or not the degree of disability caused by a given stroke merits treatment with t-PA. Score documentation my range from 0 to 42. Allowable Values: Yes or No/UTD.

• Initial NIHSS Score Time - The time (military time) for which the NIHSS score was first performed at the hospital.

Patients are eligible for the numerator population when the ICD-10-PCS Principal or Other Procedure Date and ICD-10-PCS Principal or Other Procedure Time minus the Initial NIHSS Score Date and Initial NIHSS Score Time are greater than or equal to zero minutes, OR the Initial NIHSS Score Date and Initial NIHSS Score Time minus the Arrival Date and Arrival Time are greater than or equal to zero minutes and less than or equal to 720 minutes.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Ischemic stroke patients who arrive at this hospital emergency department (ED).

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Included Populations:
• Discharges with ICD-10-CM Principal Diagnosis Code for ischemic stroke as defined in Appendix A, Table 8.1

11 data elements are used to calculate the denominator. Data elements and definitions:

• Admission Date: The month, day, and year of admission to acute inpatient care.

• Birthdate: The month, day, and year the patient was born.

• Comfort Measures Only: Comfort Measures Only refers to medical treatment of a dying person where the natural dying process is permitted to occur while assuring maximum comfort. It includes attention to the psychological and spiritual needs of the patient and support for both the dying patient and the patient´s family. Comfort Measures Only is commonly referred to as "comfort care" by the general public. It is not equivalent to a physician order to withhold emergency resuscitative measures such as Do Not Resuscitate (DNR). Allowable Values: 1 (Day 0 or 1); 2 (Day 2 or after); 3 (Timing unclear); 4 (Not documented/UTD).

• Direct Admission – Documentation that the patient was transferred from another acute care facility and taken directly to the operating room or interventional suite prior to hospital admission, or admitted directly to intensive care or other unit of the hospital. Allowable Values: Yes or No/UTD.

• Discharge Date - The month day and year the patient was discharged from acute care, left against medical advice or expired during the stay.

• Discharge Time – The documented time (military time) the patient was discharged from acute care, left against medical advice or expired during the stay.

• ED Patient - Documentation that the patient received care in a dedicated emergency department of the facility. Allowable Values: Yes or No/UTD.

• Elective Carotid Intervention - Documentation demonstrates that the current admission is solely for performance of an elective carotid intervention (e.g., elective carotid endarterectomy, angioplasty, carotid stenting). Allowable Values: Yes or No/UTD.

• ICD-10-PCS Other Procedure Codes: The International Classification of Diseases, Tenth Revision, Master Code Table (ICD-10-PCS) codes identifying all significant procedures other than the principal procedure.

• ICD-10-CM Principal Diagnosis Code: The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.

• ICD-10-CM Principal Procedure Code: The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code that identifies the principal procedure performed during this hospitalization. The principal procedure is the procedure performed for definitive treatment rather than diagnostic or exploratory purposes, or which is necessary to take care of a complication.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
• Patients less than 18 years of age
• Patients who have a Length of Stay greater than 120 days

• Patients with Comfort Measures Only documented on the day of or day after hospital arrival
• Patients admitted for Elective Carotid Intervention
• Patients who do not undergo recanalization therapy and are discharged within 12 hours of arrival at this hospital

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
• Patients less than 18 years of age.
o Patient age (in years) equals Admission Date minus Birthdate.

• Patients who have a Length of Stay greater than 120 days.
o Length of Stay (in days) equals Discharge Date minus Admission Date.

• Patients with Comfort Measures Only documented on the day of or day after hospital arrival:
o Physician/APN/PA documentation of comfort measures only (hospice, comfort care, etc.) when the earliest day of documented CMO was on the day of arrival (Day 0) or Day after arrival (Day 1).

• Patients admitted for Elective Carotid Intervention:
o Elective Carotid Intervention includes procedures of the head and neck as defined in Appendix A, Table 8.3 Carotid Intervention Procedures when medical record documentation also states that the reason for the patient's admission to the hospital was for the performance of that procedure and not for the treatment of acute ischemic stroke.
o An elective admission is documented as a pre-planned or scheduled admission to the hospital.

• Patients who do not undergo recanalization therapy and are discharged within 12 hours of hospital arrival.
o Within 12 hours of hospital arrival equals Discharge Date and Discharge Time minus Arrival Date and Arrival Time for patients who do not have an ICD-10-PCS Principal or Other Procedure Code as defined in Appendix A, Table 8.1a Thrombolytic Agent Procedures or Table 8.1b Mechanical Endovascular Reperfusion Therapy Procedures.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not Applicable

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not Applicable

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not Applicable

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*

Comprehensive Stroke (CSTK) Initial Patient Population Algorithm

Variable Key: Patient Age, Initial Patient Population Reject Case Flag, Length of Stay, Sub-Population 1 Flag, Sub-Population 2 Flag, and Sub-Population 3 Flag.

1. Start CSTK Initial Patient Population logic sub-routine. Process all cases that have successfully reached the point in the Transmission Data Processing Flow: Clinical which calls this Initial Patient Population Algorithm. Do not process cases that have been rejected before this point in the Transmission Data Processing Flow: Clinical.

2. Check ICD-10-CM Principal Diagnosis Code
a. If the ICD-10-CM Principal Diagnosis Code is not on Table 8.1 and 8.2, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the ICD-10-CM Principal Diagnosis Code is on Table 8.1 or 8.2, continue processing and proceed to the Patient Age calculation.

3. Calculate Patient Age. Patient Age, in years, is equal to the Admission Date minus the Birthdate. Use the month and day portion of admission date and birthdate to yield the most accurate age.

4. Check Patient Age
a. If the Patient Age is less than 18 years, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the Patient Age is greater than or equal to 18 years, continue processing and proceed to Length of Stay Calculation.

5. Calculate the Length of Stay. Length of Stay, in days, is equal to the Discharge Date minus the Admission Date.

6. Check Length of Stay
a. If the Length of Stay is greater than 120 days, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the Length of Stay is less than or equal to 120 days, the patient is in the CSTK Initial Patient Population.

7. Set the Initial Patient Population Reject Case Flag to equal No. Continue processing and proceed to the ICD-10-CM Principal Diagnosis Code to determine the CSTK sub-population.

8. Initialize Sub-Population 1 Flag, Sub-Population 2 Flag and Sub-Population 3 Flag to No.

9. Check ICD-10-CM Principal Diagnosis Code
a. If the ICD-10-CM Principal Diagnosis Code is on 8.2, the patient is in the CSTK Sub-Population 3 and is eligible to be sampled for the CSTK Sub-Population 3. Set Sub-Population 3 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the ICD-10-CM Principal Diagnosis Code is on 8.1, continue processing and proceed to ICD-10-PCS Principal Or Other Procedure Codes.
i. If at least one ICD-10-PCS Principal or Other Procedure Codes is on Table 8.1a or 8.1b, the patient is in the CSTK Sub-Population 2 and is eligible to be sampled for the CSTK Sub-Population 2. Set Sub-Population 2 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
ii. If none of the ICD-10-PCS Principal Or Other Procedure Codes are on Table 8.1a or 8.1b, the patient is in the CSTK Sub-Population 1 and is eligible to be sampled for the CSTK Sub-Population 1. Set Sub-Population 1 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

CSTK-01: National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients

Numerator: Ischemic stroke patients for whom a NIHSS score is performed prior to any acute recanalization therapy in patients

undergoing recanalization therapy and documented in the medical record, OR documented within 12 hours of hospital arrival for patients who do not undergo recanalization therapy.

Denominator: Ischemic stroke patients who arrive at this hospital emergency department (ED)

Variable Key: Timing I, Timing II, Timing III

1. Start processing. Run cases that are included in the Comprehensive Stroke (CSTK) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check ICD-10-CM Principal Diagnosis Code
a. If ICD-10-CM Principal Diagnosis Code is not on Table 8.1, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
b. If ICD-10-CM Principal Diagnosis Code is on Table 8.1, continue processing and proceed to ED Patient.

3. Check ED patient
a. If ED Patient is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If ED Patient equals No, continue processing and proceed to step 4 to check Direct Admission.
c. If ED Patient equals Yes, continue processing and proceed to step 5 to check Comfort Measures Only.

4. Check Direct Admission
a. If Direct Admission is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Direct Admission equals No, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If Direct Admission equals Yes, continue processing and proceed to Comfort Measures Only.

5. Check Comfort Measures Only
a. If Comfort Measures Only is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Comfort Measures Only equals 1, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If Comfort Measures Only equals 2, 3, or 4, continue processing and proceed to Elective Carotid Intervention.

6. Check Elective Carotid Intervention
a. If Elective Carotid Intervention is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Elective Carotid Intervention equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If Elective Carotid Intervention equals No, continue processing and proceed to Initial NIHSS Score Performed.

7. Check Initial NIHSS Score Performed
a. If Initial NIHSS Score Performed is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Initial NIHSS Score Performed equals No, the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Initial NIHSS Score Performed equals Yes, continue processing and proceed to Initial NIHSS Score Date.

8. Check Initial NIHSS Score Date
a. If Initial NIHSS Score Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Initial NIHSS Score Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Initial NIHSS Score Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to Initial NIHSS Score Time.

9. Check Initial NIHSS Score Time
a. If Initial NIHSS Score Time is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop

processing.
b. If Initial NIHSS Score Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Initial NIHSS Score Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to ICD-10-PCS Principal or Other Procedure Codes.

10. Check ICD-10-PCS Principal or Other Procedure Codes
a. If all missing or none ICD-10-PCS Principal or Other Procedure Codes is on Table 8.1a or 8.1b, continue processing and proceed to step 14 and check Discharge Date.
b. If any ICD-10-PCS Principal or Other Procedure Codes is on Table 8.1a or 8.1b, continue processing and proceed to ICD-10-PCS Principal or Other Procedure Code Date.

11. Check ICD-10-PCS Principal or Other Procedure Code Date
a. If ICD-10-PCS Principal or Other Procedure Code Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If ICD-10-PCS Principal or Other Procedure Code Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If ICD-10-PCS Principal or Other Procedure Code Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to ICD-10-PCS Principal or Other Procedure Code Time.

12. Check ICD-10-PCS Principal or Other Procedure Code Time
a. If ICD-10-PCS Principal or Other Procedure Code Time is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If ICD-10-PCS Principal or Other Procedure Code Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If ICD-10-PCS Principal or Other Procedure Code Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to the Timing I calculation.

13. Calculate Timing I. Timing I, in minutes, is equal to ICD-10-PCS Principal or Other Procedure Code Date and ICD-10-PCS Principal or Other Procedure Code Time minus the Initial NIHSS Score Date and Initial NIHSS Score Time.
a. If the time in minutes is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If the time in minutes is less than zero, the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If the time in minutes is greater than or equal to zero, the case will proceed to a Measure Category Assignment of E and will be in the numerator population. Stop processing.

14. Check Discharge Date
a. If Discharge Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Discharge Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Discharge Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to Discharge Time.

15. Check Discharge Time
a. If Discharge Time is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Discharge Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Discharge Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to Arrival Date.

16. Check Arrival Date
a. If Arrival Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Arrival Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Arrival Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to Arrival Time.

17. Check Arrival Time
a. If Arrival Time is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.

b. If Arrival Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Arrival Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to the Timing II calculation.

18. Calculate Timing II. Timing II, in minutes, is equal to the Discharge Date and the Discharge Time minus the Arrival Date and Arrival Time.
a. If the time in minutes is less than zero, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If the time in minutes is greater than or equal to zero and less than 720, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If the time in minutes is greater than or equal to 720, continue processing and proceed to the Timing III calculation.

19. Calculate Timing III. Timing III, in minutes, is equal to the Initial NIHSS Score Date and the Initial NIHSS Score Time minus the Arrival Date and Arrival Time.
a. If the time in minutes less than zero, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If the time in minutes is greater than 720, the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If the time in minutes is greater than or equal to zero and less than or equal to 720, the case will proceed to a Measure Category Assignment of E and will be in the numerator population. Stop processing.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter for the measure set cannot sample. Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

Two sub-populations make-up the Initial Patient Population for the CSTK-01 measure. The CSTK 1-Ischemic Stroke Without Procedure sub-population sampling group includes patients admitted to the hospital for inpatient acute care if they have: an ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.1, a Patient Age (Admission Date – Birthdate) > 18 years and a Length of Stay (Discharge Date - Admission Date) = 120 days. The CSTK 2-Ischemic Stroke with IV t-PA, IA t-PA, or MER sub-population sampling group includes patients admitted to the hospital for inpatient acute care if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.1 AND ICD-10-PCS Principal or Other Procedure Codes as defined in Appendix A, Table 8.1a OR Table 8.1b, a Patient Age (Admission Date – Birthdate) > 18 years and a Length of Stay (Discharge Date - Admission Date) less than or equal to 120 days. Both sampling groups must be sampled to meet the minimum sampling requirement for CSTK-01.

Quarterly Sampling
The Quarterly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for both CSTK sub-population 1 and sub-population 2. Hospitals performing quarterly sampling for the CSTK-01 measure must ensure that its Initial Patient Population and sample size meet the following conditions for each sub-population sampling group:
Example:
o A hospital's ischemic stroke patient population size is 200 patients during the second quarter. Fifty (50) ischemic stroke patients had a procedure for thrombolysis or mechanical clot removal. The required quarterly sample size for the CSTK-01 measure is a minimum of 84 cases (42 cases from Table 1 plus 42 cases from Table 2 equals 84).

Quarterly Sample Size Based on CSTK Sub-population 1 for Ischemic Stroke (Table 1):
Sub-Population 1: If "N" > 420, then 'n' 84
Minimum Required Sample Size: 84 records

Sub-Population 1: If “N” 211-419, then ‘n’ 20%
Minimum Required Sample Size: 20% of Sub-Population records

Sub-Population 1: If “N” 43-210, then ‘n’ 42
Minimum Required Sample Size: 42 records

Sub-Population 1: If “N” < 42, then ‘n’ 100%
Minimum Required Sample Size: No sampling; 100% Sub-Population records required

Quarterly Sample Size Based on CSTK Sub-population 2 for Ischemic Stroke with IV t-PA, IA t-PA,or MER (Table 2):
Sub-Population 2: If “N” > 420, then ‘n’ 84
Minimum Required Sample Size: 84 records

Sub-Population 2: If “N” 211-419, then ‘n’ 20%
Minimum Required Sample Size: 20% of Sub-Population records

Sub-Population 2: If “N” 43-210, then ‘n’ 42
Minimum Required Sample Size: 42 records

Sub-Population 2: If “N” < 42, then ‘n’ 100%
Minimum Required Sample Size: No sampling; 100% Sub-Population records required

Monthly Sampling
The Monthly Sample Size “n”, i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size “N” for both CSTK sub-population 1 and sub-population 2. Hospitals performing monthly sampling for the CSTK-01 measure must ensure that its Initial Patient Population and sample size meet the following conditions for each sub-population sampling group:
Example:
o A hospital’s ischemic stroke patient population size is 200 patients during March. Twenty (20) ischemic stroke patients had a procedure for thrombolysis or mechanical clot removal. The required sample size for the CSTK-01 measure is a minimum of 42 cases for the month (28 cases from Table 4 plus 14 cases from Table 5 equals 42).

Monthly Sample Size Based on CSTK Sub-population 1 for Ischemic Stroke (Table 4):
Sub-Population 1: If “N” > 140, then ‘n’ 28
Minimum Required Sample Size: 28 records

Sub-Population 1: If “N” 71-140, then ‘n’ 20%
Minimum Required Sample Size: 20% of Sub-Population records

Sub-Population 1: If “N” 15-70, then ‘n’ 14
Minimum Required Sample Size: 14 records

Sub-Population 1: If “N” < 14, then ‘n’ 100%
Minimum Required Sample Size: No sampling; 100% Sub-Population records required

Monthly Sample Size Based on CSTK Sub-population 2 for Ischemic Stroke with IV t-PA, IA t-PA,or MER (Table 5):
Sub-Population 2: If “N” > 140, then ‘n’ 28
Minimum Required Sample Size: 28 records

Sub-Population 2: If “N” 71-140, then ‘n’ 20%
Minimum Required Sample Size: 20% of Sub-Population records

Sub-Population 2: If “N” 15-70, then ‘n’ 14
Minimum Required Sample Size: 14 records

Sub-Population 2: If "N" < 14, then 'n' 100%
Minimum Required Sample Size: No sampling; 100% Sub-Population records required

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not Applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
A web-based data collection tool was developed by The Joint Commission for the pilot test process. Currently, hospitals have the flexibility of creating their own tool modeled after the pilot tool or they may develop their own data collection tools using the data element dictionary and allowable values specified in the implementation guide. Hospitals also have the option of selecting a vendor-developed data collection tool which has been verified by The Joint Commission.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not Applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2864_MeasureTesting_MSF6.5.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

Measure Title: CSTK-01: National Institutes of Health Stroke Scale Score (NIHSS) Performed for Ischemic Stroke Patients

Date of Submission: 1/15/2016

Type of Measure: Process

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions.
15. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.
16. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Not Applicable

1.3. What are the dates of the data used in testing? October 1, 2012 –March 31, 2013; first and second quarter 2015.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Hospitals were recruited for the pilot test of the measures via an open call on The Joint Commission web site. An announcement of the call with a link to The Joint Commission web site was also posted on the American Heart Association/American Stroke Association web site. Hospitals were selected with the intent to capture variability related to ownership, size, type of facility and location. Eighty-two hospitals from twenty-seven states were selected from more than 120 volunteers to participate in the six-month pilot test of the measures. Twenty hospitals withdrew during the pilot test citing lack of resources to complete the project. Sixty-two hospitals submitted data for each month of the six-month pilot test. An additional four hospitals submitted data for one or more months.

Sixty-six hospitals contributed data for the analysis of the measures:

Ownership:

For Profit	14
Not for Profit	52

Bed Size:	
Less than 100	0
100 – 199	3
200 – 299	7
300 – 499	26
Greater than 500+	30

Located in 27 states:
Alabama
Arizona
California
Colorado
Florida
Georgia
Illinois
Indiana
Kentucky
Louisiana
Maryland
Massachusetts
Michigan
Minnesota
Missouri
Nevada
New Jersey
New York
North Carolina
Ohio
Oregon
Pennsylvania
Tennessee
Texas
West Virginia
Washington
Wisconsin

Other:	
Teaching	41
Non-teaching	25

Urban	59
Rural	7

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

During the six-month pilot test, sixty-six hospitals submitted data for 4,686 inpatient records. The cases included patients greater than 18 years of age, male and female, all races, and all payers. Age, gender, racial, and payer distribution are not known because the results were de-identified. The patient population was comprised of all patients discharged with an ICD-9-CM Principal Diagnosis Code for ischemic stroke.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Data from the six month pilot test of the measures were used for reliability testing and face validity. To test the empirical validity of the measures, two quarters of data from 42 Joint Commission certified comprehensive stroke centers were used to conduct a secondary analysis.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Reliability testing was performed at twelve participating pilot sites. Testing was conducted on a stratified random sample of records selected from each organization at the organization and measure category level. Hospitals were visited by teams of two Joint Commission staff during April, May, June, July, and August 2013. During these visits, Joint Commission staff re-abstracted records previously abstracted by hospital staff. A total of 281 records were re-abstracted. In cases of disagreement between the re-abstracted and original abstraction on a data element, reasons for disagreement were determined and adjudication was made as to whether original or re-abstraction findings were correct. Reliability was addressed by comparing the original abstracted and adjudicated re-abstracted values, with the adjudicated value serving as the gold standard. The data analysis included both the percent agreement and the kappa statistic to adjust for chance agreement for categorical data elements.

2a2.3. For each level checked above, what were the statistical results from reliability testing? (e.g., *percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data Element	Number of Mismatches	Match Rate	Kappa
Arrival Date	7	97.5%	NA
Arrival Time	50	82.2%	NA
ICD-9-CM Other Procedure Dates	15	94.1%	NA
ICD-9-CM Other Procedure Times	39	86.0%	NA
ICD-9-CM Principal Procedure Date	15	94.1%	NA
ICD-9-CM Principal Procedure Time	15	94.1%	NA
Initial NIHSS Score Date	18	93.6%	NA
Initial NIHSS Score Performed	6	97.9%	0.95
Initial NIHSS Score Time	52	81.5%	NA
Discharge Date	9	96.8%	NA
Discharge Time	80	71.5%	NA

ED Patient	2	99.3%	0.96
Elective Carotid Intervention	3	98.9%	-0.0043
Warning Signs and Symptoms of Stroke	20	89.3%	0.26

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i*.e., what do the results mean and what are the norms for the test conducted?*)

A statistical measure of inter-rater reliability is Cohen's Kappa, which ranges generally from 0.0 to 1.0 (although negative numbers are possible), where large values mean better reliability and values near zero suggest that agreement is attributable to chance alone. It indicates the proportion of agreement not expected by chance alone (e.g., Kappa of 0.6 means that raters agreed 60% of the time over and above what would be expected by chance alone).

Landis & Koch, 1977 offers the following clarification of Kappa interpretation:

< 0	Poor agreement
0.00-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-1.00	Almost perfect agreement

Other authors (Cicchetti & Sparrow; Fleiss) have suggested additional classifications for interpreting the Kappa statistic, but all seem to indicate Kappa > 0.60 is desirable.

The statistical measure of inter-rater reliability is suitable for the categorized data type, therefore the measure is applied to the following data elements: *Initial NIHSS Score Performed*, *ED Patient*, *Elective Carotid Intervention*, and *Warning Signs and Symptoms of Stroke*. The Kappa values indicate almost perfect agreement for the data elements *NIHSS Score Performed* and *ED Patient*; the Kappa value indicates fair agreement for the *Warning Signs and Symptoms of Stroke* data element; even though the Kappa value indicates poor agreement to the *Elective Carotid Intervention*, the low marginal proportion of those with this intervention in the sample (<1%) means that Kappa in this case is not well determined. According to the statistical measure of Kappa value, we believe these results demonstrate acceptable reliability of the assessment data used in the performance measure.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

X **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

X **Empirical validity testing**

X **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Face Validity:

Measure face validity was assessed via survey and focus groups of hospitals participating in the pilot test. Focus group discussions were held at all test sites visited, during which we received feedback as to whether the measure, data elements, and definitions accurately reflected existing evidence. All of the respondents indicated that all aspects of the measures accurately reflected current evidence. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant.

To determine feasibility and identify areas for potential revision, test sites were asked to electronically rate the clarity of numerator statements, denominator statements, and measure information forms (MIFs) on a five point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree). Data elements and associated tables were evaluated for clarity, accuracy, data availability and accessibility.

Empirical Validity:
Measure convergent validity was assessed using hospitals patient level data from The Joint Commission warehouse. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous face validity testing. We conducted a secondary analysis to help interpret results; correlation among CSTK process measures.

The data were comprised of first and second quarter 2015 submissions. This included 42 hospitals submitting 65,389 inpatient records for the selected CSTK measures CSTK-1, CSTK-2, CSTK-3 and CSTK-6. The hospital's selection was based on those hospitals that reported 6 months of data and had 30 or more denominator cases for the period.

Comprehensive Stroke (CSTK) Initial Patient Population
The CSTK Initial Patient Population is unique in that it is comprised of three distinct subpopulations: ischemic stroke patients who do not undergo a reperfusion therapy (i.e., procedure), ischemic stroke patients who undergo a reperfusion therapy (IV t-PA, IA t-PA, or mechanical endovascular reperfusion (MER) therapy), and hemorrhagic stroke patients.

Subpopulation 1: Ischemic Stroke Without Procedure
This subpopulation comprises ischemic stroke patients who are admitted to the hospital for inpatient acute care and do not undergo a reperfusion procedure (CSTK-01 measure). Patients are included in the CSTK-1 Ischemic Stroke Without Procedure subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.1, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

Subpopulation 2: Ischemic Stroke With IV t-PA, IA t-PA, or MER
This subpopulation comprises ischemic stroke patients who receive IV t-PA, IA t-PA, or MER procedures during the hospital stay (CSTK-01 and CSTK-02 measures). Patients admitted to the hospital for inpatient acute care are included in the CSTK-2 Ischemic Stroke With IV t-PA, IA t-PA, or MER subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.1 and ICD-10-PCS Principal or Other Procedure Codes as defined in Appendix A, Table 8.1a OR Table 8.1b, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

Subpopulation 3: Hemorrhagic Stroke
This subpopulation comprises hemorrhagic stroke patients admitted to the hospital for inpatient acute care (CSTK-03 and CSTK-06 measures). Patients are included in the CSTK-3-Hemorrhagic Stroke subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.2, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Pilot Test Data:
Descriptive statistics for the measure: N= 66 pilot site hospitals
Overall rate=67% (SD=26%), min=0%, max=100%, median=70.8%

Additional analyses:
CSTK01 and CSTK02 are measures for ischemic stroke. The measure rate among these measures are expected to be correlated. The Pearson correlation coefficient is calculated.

Pearson Correlation Coefficient interpretation: the range is from -1 to 1. When the correlation coefficient is close to +1 or -1, it means that there is strong correlation; p value is utilized to determine if the corrlation coefficient is significant or not. If it is less than 0.05, then the conclusion usually is significant.

The table below demonstrates that the CSTK01 rate was positively correlated to the CSTK-02 measure. Because the data is not normally distributed, the correlation coefficient on correlation may not be the best measure of association.

Pearson Correlation Coefficients, N = 66 Prob > |r| under H0: Rho=0

	CSTK01	**CSTK02**
Correlation Coefficient	1.00000	0.25543
P value		0.0385

The average rating for measure CSTK-01 numerator and denominator statements, including the clarity of numerator and denominator inclusions and exclusions was 4.37, indicating that the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.
Data element validity was evaluated for clarity and accuracy, as well as, data availability and accessibility. The percentage of agreement for new data elements developed specifically for CSTK-01 and other measures in the CSTK measure set is detailed in the table below:

Data Element Name	Clarity / Accuracy	Availability / Accessibility
ICD-9-CM Other Procedure Times	90.91%	47.73%
ICD-9-CM Principal Procedure Time	100.00%	47.73%
Initial NIHSS Score Date	100.00%	47.73%
Initial NIHSS Score Performed	95.45%	47.73%
Initial NIHSS Score Time	95.45%	47.73%
Discharge Time	95.45%	47.73%
Warning Signs and Symptoms of Stroke	97.67%	47.73%

1Q and 2Q 2015 Data:
Overall descriptive statistics for CSTK selected measures: N = 42 certified comprehensive stroke hospitals: n = 65,389

CSTK-01
Median: 89%
Percentile 10%: 70%
Percentile 25%: 80%
Percentile 75%: 93%
Percentile 90%: 95%

CSTK-02
Median: 95%
Percentile 10%: 63%
Percentile 25%: 75%
Percentile 75%: 100%
Percentile 90%: 100%

CSTK-03
Median: 61%

Percentile 10%: 33%
Percentile 25%: 48%
Percentile 75%: 79%
Percentile 90%: 89%

CSTK-06
Median: 86%
Percentile 10%: 74%
Percentile 25%: 80%
Percentile 75%: 93%
Percentile 90%: 100%

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
CSTK_1	42	0.83687	0.17097	35.14857	0.00267	1.00000
CSTK_2	41	0.85894	0.20516	35.21648	0.20000	1.00000
CSTK_3	42	0.59308	0.22979	24.90933	0.07273	0.92969
CSTK_6	42	0.83555	0.17123	35.09314	0	1.00000

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	CSTK_1	CSTK_2	CSTK_3	CSTK_6
CSTK_1	1.00000 42	0.41205 0.0074 41	0.65340 <.0001 42	0.69186 <.0001 42
CSTK_2	0.41205 0.0074 41	1.00000 41	0.28658 0.0693 41	0.31724 0.0433 41
CSTK_3	0.65340 <.0001 42	0.28658 0.0693 41	1.00000 42	0.55910 0.0001 42
CSTK_6	0.69186 <.0001 42	0.31724 0.0433 41	0.55910 0.0001 42	1.00000 42

The CSTK measures table shows a positive correlation and statistical significance which indicates that hospitals with high quality on one CSTK measure tend to have high correlations on the other stroke measures.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)
Pilot Test Data:

We note that ratings for both data elements and the measure are relatively high; numerators, denominators and measures were ranked above the midpoint of 3.0 (average), and data elements were above 75% positive in clarity, collectability, and correctness of data sources. We conclude that the measures and specifications are valid.

1Q and 2Q 2015 Data:
Overall the positive inter-correlations indicate convergent validity of all the measures.
They are positively correlated with other evidence-based processes of care.

2b3. EXCLUSIONS ANALYSIS .

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Pilot Test Data:
The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patients who have a Length of Stay greater than 120 days
- Patients admitted for Elective Carotid Intervention
- Patients who do not undergo re-canalization therapy and are discharged within 12 hours of arrival at this hospital
- Patients without warning signs and symptoms of stroke on arrival at this hospital

1Q and 2Q 2015 Data:
The following exclusions were analyzed by subpopulation and measure for frequency and variability across providers:

- Patients less than or equal to 18 years of age
- Patients who have a Length of Stay greater than or equal to 120 days
- *Patients with Comfort Measures Only* documented on the day of or day after hospital arrival
- *Patients admitted for Elective Carotid Intervention*
- *Patient*s who do not undergo recanalization therapy and are discharged within 12 hours of arrival at time

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Pilot Test Data:
There were 10668 admissions included in the initial cohort and diagnosed ischemic stroke. From among the 10668 admissions in 66 hospitals, the descriptive statistics are given below.

Exclusion: Patient not in CSTK Initial Patient Population
Note: A case was excluded from the Initial Patient Population as determined by the following:

- Patients less than 18 years of age: Overall Occurrence n = 39 (0.37%)
- Patients who have a Length of Stay greater than 120 days: Overall Occurrence n = 12 (0.11%)

There were 6508 admissions included in the initial cohort and diagnosed ischemic stroke. From among the 6508 admissions in 66 hospitals, the descriptive statistics are given below.

Exclusion: Elective Carotid Intervention
Overall Occurrence n = 22
Overall Occurrence Percentage: 0.34%

Exclusion: Patients who do not undergo re-canalization therapy and are discharged within 12 hours of arrival at this hospital
Overall Occurrence n = 9
Overall Occurrence Percentage: 0.14%

Exclusion: Patients without warning signs and symptoms of stroke on arrival at this hospital
Overall Occurrence n = 378
Overall Occurrence Percentage: 5.81%

1Q and 2Q 2015 Data:
There were 65,389 admissions selected from the initial cohort. From among the 65,389 admissions in 42 hospitals, the descriptive statistics are given below.

Applied To Measure CSTK-01 CSTK-03 CSTK-06
Exclusion: *Comfort Measures* - 1 Day 0 or 1:
Overall Occurrence n = 1,300
Overall Occurrence Percentage: 2%
Minimum: 0.47%
Median: 3%
Maximum: 6%

Applied To Measure CSTK-01 CSTK-02
Exclusion: Patients admitted for *Elective Carotid Intervention*
Overall Occurrence n = 602
Overall Occurrence Percentage: 2%
Minimum: 0.24%
Median: 3.0%
Maximum: 7.64%

Applied To Measure CSTK-01
Exclusion: Patients who do not undergo recanalization therapy and are discharged within 12 hours of arrival at time
Overall Occurrence n = 10,181
Overall Occurrence Percentage 63%
Minimum: 47%
Median: 53%
Maximum: 100%
NMISS= 10,125 missing observation

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Pilot Test Data:
According to the overall occurrences in 2b3.2, the overall frequency of exclusions is low for those in the measure denominator. The distribution of exclusions across hospitals is very narrow indicating that the occurrence is random and likely would not bias performance results.

It is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age

Rationale: The literature does not support use in the pediatric patient.

Patients who have a Length of Stay greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Patients admitted for Elective Carotid Intervention
Rationale: Patients excluded because the purpose of admission to the hospital was for the performance of an intervention to prevent stroke.

Patients who do not undergo recanalization therapy and are discharged within 12 hours of arrival at this hospital
Rationale: It is inappropriate to include patients who do not undergo recanalization therapy whose hospital stay is less than the timeframe for performing NIHSS.

Patients without warning signs and symptoms of stroke on arrival at this hospital
Rationale: It is inappropriate to include patients who do not present to the hospital with stroke symptoms.

1Q and 2Q 2015 Data:
The overall frequency of exclusions is low for those in the measure denominator. The distribution of exclusions across hospitals is narrow indicating that the occurrence is random and likely would not bias performance results.

It is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a *Length of Stay* greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Patients with *Comfort Measures Only* documented
Rationale: It is inappropriate to treat such patients beyond those measures necessary for comfort.

Patients admitted for *Elective Carotid Intervention*
Rationale: Patients are excluded because the purpose of admission to the hospital was for the performance of an intervention to prevent stroke.

Patients who do not undergo recanalization therapy and are discharged within 12 hours of arrival at this hospital
Rationale: It is inappropriate to include patients who do not undergo recanalization therapy whose hospital stay is less than the timeframe for performing NIHSS.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

Not Applicable

2b4.1. What method of controlling for differences in case mix is used?
X **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
Not Applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care and not related to disparities*)
Not Applicable

2b4.4. What were the statistical results of the analyses used to select risk factors?
Not Applicable

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)
Not Applicable

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
if stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
Not Applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Not Applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Not Applicable

2b4.9. Results of Risk Stratification Analysis:
Not Applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)
Not Applicable

***2b4.11. Optional Additional Testing for Risk Adjustment** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*
Descriptive statistics for the performance measure scores for all tested entities were constructed. These statistics were the mean, standard deviation, median, minimum, and maximum scores.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

A meaningful difference was defined as a spread of >30 percentage points between the median and maximum pilot hospital rates.

Results for CSTK-01 were:
Descriptive statistics for measure: N=66 hospitals
Overall rate=67%
Standard Deviation=26%
Minimum=0%
Median=70.8**%**
Maximum=100%

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

The results are interpreted as showing a meaningful spread between both the minimum and median scores and between minimum and maximum scores. This is indicative of statistically significant differences.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
***If only one set of specifications, this section can be skipped*.**

Not Applicable

Note: *This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different datasources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not Applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not Applicable

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i*.e., what do the results mean and what are the norms for the test conducted*)

Not Applicable

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other

If other: Data element allowable values are selected, either manually or electronically, from clinical and coded data available in medical record documentation. All medical record documentation is used in the abstraction process. Vendor data collection tools are used to import data elements needed for measure rate calculation.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract this measure electronically, so offers a chart-abstracted version which allows for data capture from unstructured data fields. The Joint Commission plans to retool the measure for capture from electronic sources within the next several years.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

This measure was collected during a 6 month pilot process for Joint Commission Comprehensive Stroke Certification. It was learned via discussions and written pilot evaluation materials that there were some issues related to data abstraction of the following data element:

- Initial NIHSS Score Performed
  - Clarification was requested regarding the use of modified NIHSS scores and scores estimated from documented components in the absence of a total documented NIHSS score.

These issues related to data abstraction have been easily resolved through clarification of guidelines for abstraction.

Based on feedback from the pilot sites, the measure algorithm was modified to add the following data elements:

- Direct Admission

o Addition of this data element was requested to capture in the denominator population those stroke patients transferred from another acute care facility and taken directly to the interventional suite or operating room, bypassing the emergency department prior to hospital admission.

- Comfort Measures Only

o Addition of this data element was requested to exclude stroke patients who are Comfort Measures Only on the day of or day after hospital arrival.

The measure algorithm was further modified to remove the following data element:

- Warning Signs of Symptoms of Stroke

o Removal of this data element was requested because it was redundant with the principal diagnosis of ischemic stroke, consistently abstracted "Yes", and increased abstraction burden.

Other information impacting the feasibility and implementation of the measure was also obtained from the pilot process and is summarized as follows:

Staff Training and Education:
To prepare for and support continuous data collection throughout the pilot test, a total of ten hours was spent on staff training and education. Training was accomplished via two 2-hour webinars and monthly conference calls with pilot site participants.

Case Identification/Medical Record Retrieval:
Case identification was not a problem; cases were identified by the ICD-9-CM principal diagnosis codes for ischemic stroke. Record retrieval time varied depending on the type of medical record. On average, 10 minutes were spent for record retrieval with more time spent to retrieve a paper record than electronic health record.

Case Selection:
For the pilot test of the measure, sampling was allowed. The sampling methodology was modified post-pilot to balance the inclusion of ischemic stroke patients who do not undergo recanalization therapy and those ischemic stroke patients who receive IV or IA thrombolytic therapy or mechanical reperfusion therapy in the denominator population.

Data Abstraction:
Medical record review to abstract all data elements required for the measure set averaged 45 minutes per record. Time spent on record review varied with case complexity and the number of procedures and interventions performed, as well as, the number of data elements collected for the measure. In general, ischemic stroke cases required more time for review than did hemorrhagic stroke cases.

Data abstraction was primarily done by nurses, (e.g., Registered Nurse(s) with a quality improvement background, Stroke Coordinators, and Advanced Practice Nurses). Some pilot sites reported that the abstractor reviewed the record with the medical director or neurologist at least initially to identify documentation of measure specific data elements. Data specialists or administrative staff were utilized to enter abstracted data into the on-line data collection tool.

Cost of Data Abstraction:
Using 2012 national wage averages, it is estimated that the cost per case to abstract for this measure was approximately $3.50

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
This measure is one in a set of measures implemented with discharges effective January 1, 2015 for Joint Commission Comprehensive Stroke Certification. There are no fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Public Health/Disease Surveillance Paul Coverdell National Acute Stroke Registry http://www.cdc.gov/dhdsp/programs/stroke_registry.htm Quality Improvement (Internal to the specific organization) Disease-Specific Care Certification for Comprehensive Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

• Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

• Name of program and sponsor: Paul Coverdell National Acute Stroke Registry; Centers for Disease Control and Prevention
• Purpose: A national registry that measures, tracks, and improves the quality of care and access to care for stroke patients from onset of stroke symptoms through rehabilitation and recovery; decreases rate of premature death and disability from stroke; eliminates disparities in care; supports the comprehensive stroke system across the continuum of care; improves access to rehabilitation and opportunities for recovery after stroke; and, increases the workforce capacity and scientific knowledge of stroke care within stroke systems of care
• Geographic area and number and percentage of accountable entities and patients included: 11 states; 403 hospitals (CDC, 2014)

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Not Applicable

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Not Applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Sizeable improvement has been noted since the pilot test of the measure based on the first two quarters of Joint Commission ORYX

performance measurement data. The initial performance measure gap of approximately 28% has decreased by 10-11% as demonstrated by a national aggregate rate of 83% (N=50) for 2Q2015; 4-6% gap in the upper quartile of hospitals, ~11% gap for the median, and 27% gap exists for the lowest decile of hospitals.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not Applicable

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no unintended negative consequences reported or detected during testing or since implementation of the measure specifications.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Not applicable

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Karen, Kolbusz, kkolbusz@jointcommission.org, 630-792-5931-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Ann, Watt, awatt@jointcommission.org, 630-792-5944-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel was to provide advisory oversight in the literature review, measure construct and content, review of testing results, and endorsement of draft and finalized measures. Additionally they may be called upon in the future to provide measure content oversight and updates.

Technical Advisory Panel

Harold P. Adams, Jr., MD
Professor of Neurology
Director of the Division of Cerebrovascular Diseases
Director of the Stroke Center
University of Iowa Health Care
Iowa City, IA

Mark J. Alberts, MD
Clinical Vice-Chair
Department of Neurology and Neurotherapeutics
The University of Texas Southwestern Medical Center
Dallas, TX

Anne W. Alexandrov, PhD, RN, FAAN
Health Outcomes Institute
Fountain Hills, AZ

Andrew W. Asimos, MD
Director of Emergency Stroke Care
Carolinas Medical Center
Charlotte, NC

Thomas P. Bleck, MD, FCCM
Professor of Neurological Science, Neurosurgery, Internal Medicine, and Anesthesiology
Associate Chief Medical Officer, Critical Care
Rush Medical College
Chicago, IL

Darlene Bradley, RN, PhDc, MSN, MAOM
Director Emergency/Trauma Services
University of California Irvine Medical Center
Orange, CA

Mary G. George, MD, MSPH FACAS
Medical Officer
Division for Heart Disease and Stroke Prevention
Centers for Disease Control and Prevention
Atlanta, GA

Martin Gizzi, MD, PhD
Legacy Health System
Portland, OR

James, Grotta, MD
Professor and Chairman, Department of Neurology
Chief, Vascular Neurology
University of Texas Medical School-Houston
Houston, TX

Edward C. Jauch, MD, MS
Professor, Division of Emergency Medicine and Department of Neurosciences
Medical University of South Carolina
Charleston, SC

Irene Katzan, MD, MS
Director, Neurological Institute Center for Outcomes Research and Evaluation
Cleveland Clinic
Cleveland, OH

Jawad F. Kirmani, MD
Director of Stroke and Cerebrovascular Program
JKF Medical Center
Edison , NJ

Brajesh K. Lal, MD
University of Maryland Medical Center,
Division of Vascular Surgery
Baltimore, MD

Kimberly E. Levasseur-Franklin, PharmD, BCPS
Critical Care Pharmacist, Neuro/Spine ICU
Clinical Leader, ICU Pharmacist Work Group
Northwestern Memorial Hospital
Chicago, IL

Barbara Lutz, PhD, RN, CRRN, APHN-BC, FAHA, FNAP, FAAN
McNeill Distinguished Professor
School of Nursing / College of Health and Human Services
University of North Carolina-Wilmington
Wilmington, NC

Cameron McDougall, MD, FRCSC
Barrow Neurological Institute
Phoenix, AZ

Kathy Morrison, RN, MSN, CNRN
Stroke Program Coordinator
Penn State Hershey Medical Center
Hershey, PA

Charles J. Prestigiacomo, MD, FAANS, FACS
Professor, Department of Neurological Surgery, Radiology, Neurology and Neurosciences
Director, Residency Program in Neurological Surgery
Director, Cerebrovascular and Endovascular Surgery
University of Medicine and Dentistry of New Jersey
Newark, NJ

Howard A. Rowley, MD
Chief of Neuroradiology
Joseph Sackett Professor of Radiology
Professor of Radiology, Neurology, and Neurosurgery
University of Wisconsin
Madison, WI

Jeffrey Saver, MD, FAHA, FAAN
Professor of Neurology
Stroke Center Director
Professor of Medicine
David Geffen School of Medicine at UCLA,
Ronald Reagan UCLA Medical Center
Los Angeles, CA

Lee H. Schwamm, MD, FAHA
Vice Chair of Neurology, Professor of Neurology
Harvard Medical School
Chair, American Heart Association GWTG Steering Committee
Director, Stroke Program/Telemedicine
Massachusetts General Hospital
Boston, MA

Deborah, Summers, MSN, ACNS-BC, CNRN, CCRN, FAHA
Stroke Program Coordinator
St. Luke's Brain and Stroke Institute
Kansas City, MO

Dileep R. Yavagal, MD
Director, Interventional Neurology
Assistant Professor, Clinical Neurology and Neurosurgery
University of Miami and Jackson Memorial Hospital
Miami, FL

Osama O. Zaidat, MD, MS, FAAN, FAHA
St. Vincent Mercy Hospital
Toledo, OH

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2015

Ad.3 **Month and Year of most recent revision:** 08, 2015

Ad.4 **What is your frequency for review/update of this measure?** biannual

Ad.5 **When is the next scheduled review/update for this measure?** 02, 2016

condition of usage: 1) disclosure that the Specifications Manual is periodically updated, and that the version being copied or reprinted may not be up-to-date when used unless the copier or printer has verified the version to be up-to-date and affirms that, and 2) users participating in Joint Commission accreditation, including ORYX® vendors, are required to update their software and associated documentation based on the published manual production timelines.

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2865
Measure Title: CSTK-02: Modified Rankin Score (mRS) at 90 Days
Measure Steward: The Joint Commission
Brief Description of Measure: Proportion of ischemic stroke patients age 18 years and older treated with intra-venous (IV) or intra-arterial (IA) thrombolytic (t-PA) therapy or who undergo mechanical endovascular reperfusion therapy for whom a 90 day (greater than or equal to 75 days and less than or equal to 105 days) mRS is obtained via telephone or in-person.

This is the second measure in a set of measures developed for Joint Commission Comprehensive Stroke Certification. The other measures in the set include CSTK-01 National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients; CSTK-03 Severity Measurement Performed for Subarachnoid Hemorrhage (SAH) and Intracerebral Hemorrhage (ICH) Patients (Overall Rate); CSTK-06 Nimodipine Treatment Administered. Although it is not required that these measures are reported in conjunction with each other, The Joint Commission develops measures in sets in order to provide as comprehensive a view of quality for a particular clinical topic as possible.
Developer Rationale: The Modified Rankin Scale (mRS) is the accepted standard for assessing recovery post-stroke. As such, it has become the most widely used clinical outcome measure for stroke clinical trials. Scores are used to measure the degree of disability or dependence in activities of daily living. Score reliability and reproducibility are improved through use of a structured interview by a trained evaluator. Interviews may be conducted in-person or over the phone (Schwamm, 2009). According to guideline recommendations from the American Heart Association/American Stroke Association, standardized interviews to obtain a mRS score should be conducted for acute ischemic stroke patients treated with IV or IA thrombolytic (t-PA) therapy or mechanical endovascular reperfusion therapy at 3 months (90 days) (Leifer, 2011); however, recovery may continue well beyond 3 months for many ischemic stroke patients.

Numerator Statement: Ischemic stroke patients for whom a 90 day (greater than or equal to 75 days and less than or equal to 105 days) mRS is obtained via telephone or in-person.
Denominator Statement: Ischemic stroke patients treated with IV or IA thrombolytic (t-PA) therapy or who undergo mechanical endovascular reperfusion therapy.
Denominator Exclusions:
- Patients less than 18 years of age
- Patients who have a Length of Stay greater than 120 days
- Patients admitted for Elective Carotid Intervention
- Patients who expire during the hospital stay

Measure Type: Process
Data Source: Electronic Clinical Data, Paper Medical Records
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☐ **Yes** ☒ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☐ **Yes** ☒ **No**
- **Evidence graded?** ☐ **Yes** ☒ **No**

Evidence Summary

- The evidence for this process measure should demonstrate how conducting an Modified Rankin Score (mRS) 90 days after acute stroke treatment relates to health outcomes. The developer's diagram does not show such a relationship. The mRS is used to assess functional outcome and the extent of disability after stroke. In the absence of use of the mRS numerical result or comparing the result to a benchmark (which would be an outcome), there is no empirical evidence presented to support this process measure.
- Though not presented in the Evidence section, the developer mentions a recommendation from AHA/ASA in the rationale, as follows: "*standardized interviews to obtain a mRS score should be conducted for acute ischemic stroke patients treated with IV or IA thrombolytic (t-PA) therapy or mechanical endovascular reperfusion therapy at 3 months (90 days)*" (Leifer, *Metrics for Measuring Quality of Care in Comprehensive Stroke Centers: Detailed Follow-Up to Brain Attack Coalition Comprehensive Stroke Center Recommendations*, 2011). This publication does not provide empirical evidence for the recommendation – thus it appears to be expert opinion only.
- A systematic review regarding the reliability of the Modified Rankin Score (mRS) is described. This evidence addresses the reliability and validity of the mRS data element in the measure, but does not provide supporting evidence for this process measure. The summary of this evidence suggests that it is of moderate quality, with somewhat inconsistent findings. The systematic review concluded that the reliability of the mRS is moderate.
- A recommendation from the 2009 Review of the Evidence for the Use of Telemedicine within Stroke Systems of Care from the American Heart Association/American Stroke Association discusses conducting the mRS evaluation by phone. The recommendation ("*Telephonic assessment for measuring functional disability after stroke is recommended…*") is a class 1 recommendation (signifying that it should be performed) with grade B evidence (meaning that it is based on limited populations and is derived from a single randomized trial or nonrandomized studies).

Exception to evidence– In the absence of empirical evidence to support this process measure, the Committee may consider an exception to the evidence requirement with adequate justification.

Guidance from the Evidence Algorithm

Process measure/no systematic review (Box 3) → no empiric evidence (Box 7) → INSUFFICIENT – Committee to determine whether an exception is justified.

Questions for the Committee:

- *Are you aware of any evidence linking conducting a functional status assessment using the mRS (without use of the numerical result) to health outcomes?*
- *For possible exception to the evidence criterion:*
  - *Are there, or could there be, performance measures of a related health outcome, OR evidence-based*

intermediate clinical outcomes, intervention/treatment?

- *Is there evidence of a systematic assessment of expert opinion beyond those involved in developing the measure?*
- *Does the SC agree that it is acceptable (or beneficial) to hold providers accountable for this process of care without empirical evidence?*

Preliminary rating for evidence: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
Rationale: Lack of empirical evidence to support the process of care

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer provided the following data for current performance:.

	Pilot test (10/2012-3/2013)	Q1 2015	Q2 2015
# sites	66	9	40
# patients (den)	1,307	17	579
Hospital average rate	27%	65%	88%
Standard deviation	30.3%	44%	19%
Hospital median	70%	100%	98%
Range	0 -100%	10th percentile = 0% 25th percentile = 33% 90th percentile = 100%	10th percentile = 58% 25th percentile = 82% 90th percentile = 100%
National aggregate rate		59%	87%

Disparities

- The developer cites several studies that examined variations in post-stroke functional status for men vs. women and for blacks vs. whites. However, these studies did not examine sub-population differences in assessing functional status post-stroke (the focus of this measure), so it is unclear whether there are disparities in this process of care.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence that disparities exist in the assessment of functional status by hospitals after stroke?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **Evidence for the impact of this measure is not provided.

**Almost no evidence that this is a measure of quality. Its very hard to imagine how measuring an mRS, per se, might improve

outcomes for patients. One can imagine how this information would be included in a measure (i.e. adjusted hospital mRS at 90 days) or that the process of obtaining an mRS (i.e. 90 day phone call) would be a process that could improve outcomes for patients, but its exceedingly hard to think how measuring function may accomplish that.

1b. Performance Gap

Comments: **I am uncomfortable describing the lack of uniform performance of the recommended measure as a "performance gab" because there is no evidence presented that compliance with this measure represents superior performance. There is variability in the performance of this measure, however no link to outcomes has been established so to refer to this variability as a "performance gap" is misleading.

**There is evidence from the pilot that there is suboptimal mRS recording.

While disparities in post-stroke outcomes are known to exist by race, I don't see any evidence in differences in mRS documentaiotn by race.

1c. High Priority (previously referred to as High Impact)

Comments: **N/A

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Medical record abstraction (electronic or paper). This is not an eMeasure.

Specifications:

- The measures is specified at the facility level of analysis, in the hospital/acute care facility setting.
- Two data elements define the numerator (mRS date and mRS score). Patients are eligible for inclusion in the numerator if the mRS is administered between 75 and 105 days post-discharge. The mRS can be obtained either in-person or via telephone.
    - Allowable values for the actual mRS score include:
        - 0—No residual symptoms
        - 1—No significant disability: able to carry out all pre-stroke activities
        - 2—Slight disability: unable to carry out all pre-stroke activities but able to look after self without daily help
        - 3—Moderate disability: requires some external help but able to walk without the assistance of another individual
        - 4—Moderately severe disability: unable to walk or attend to bodily functions without assistance of another individual
        - 5—Severe disability: bedridden, incontinent, requires continuous care
    - Other allowable values for the measure numerator include:
        - 6—Patient has died
        - 7—Unable to contact patient/caregiver
        - 8—mRS not performed OR unable to determine (UTD) from medical record documentation.
- Ten data elements define the denominator, which includes ischemic stroke patients treated with IV or IA thrombolytic (t-PA) therapy or who undergo mechanical endovascular reperfusion therapy. The denominator data elements include admission, discharge, and birth dates; discharge disposition; ICD-10 diagnosis and procedure codes; elective carotid intervention indicator; and four exclusion codes. Specific ICD-10 diagnosis and procedure codes are included in an attached spreadsheet.
- Patients are excluded from the measure if they are under age 18, have a length of stay greater than 120 days, are admitted for elective carotid intervention, or die during the hospital stay.
- A detailed calculation algorithm is provided.
- Sampling monthly or quarterly is allowed; instructions for sampling are provided.

- A web-based data collection tool has been developed but is not described. Hospitals are not required to use this tool.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- Inter-rater reliability testing of three of the data elements (the two numerator data elements and one exclusion data element) was conducted at 12 sites with 281 total records from 2013. Percent agreement and Kappa scores were used to compare two sets of abstracted data. Analysis of inter-rater reliability is an appropriate test for data element reliability for abstracted data.
- NQF guidance indicates that data element testing should be conducted for all critical data elements, although at minimum, results about the numerator, denominator, and exclusions should be provided. For this measure, the developers do not provide reliability testing results for the denominator data elements or for three of the four exclusion data elements. However, absence of testing for denominator data elements (and possibly for exclusion data elements) may be warranted if cases are identified through use of billing or other administrative data.
- As noted in the evidence section, above, the developers provided information regarding the reliability of the mRS itself (described as "moderate" based on a systematic review).

Results of reliability testing

- Percent agreement values for the three data elements examined ranged from 85.1% (mRS score) to 98.9% (elective carotid intervention). A Kappa score (kappa=-0.0043) was calculated for the elective carotid intervention variable.
- The kappa statistic—appropriate for non-continuous variables—represents the proportion of agreement between two abstractors that is not explained by chance alone. Values for kappa range between -1.0 and 1.0. A value of 1.0 reflects perfect agreement; a value of 0 reflects agreement that is no better than what would be expected by chance alone; a value less than zero reflects agreement that is worse than what would be expected by chance (and potentially, systematic disagreement). Percent agreement does not adjust for chance agreement and therefore should not be used alone to demonstrate reliability.
- Very low kappa values usually are interpreted as reflecting poor agreement. However, the kappa statistic is influenced by the prevalence of the attribute. The developers suggest that the low prevalence of patients with elective carotid intervention in the testing sample makes the value of kappa less informative and they instead rely on the percentage agreement value in their interpretation of this data element's reliability.

Guidance from the Reliability Algorithm

Precise specifications (Box 1) →empirical testing with statistical tests as specified (Box 2) → reliability testing with patient-level data elements (Box 8) →assessed all critical data elements for numerator and one of the exclusions, with the remaining denominator data elements obtained via administrative data→ high/moderate certainty the data elements are reliable (Box 10a) → Moderate

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☐ **Yes** ☐ **Somewhat** ☒ **No**

Specification not completely consistent with evidence: The specifications are consistent with an expert opinion-based recommendation but not empirical evidence.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- Empirical testing of the measure score:
  - Pilot testing data from October 1, 2012 – July 15, 2013 included information from 1,307 inpatient records from 66 hospitals. Additional testing data for the period Q1-Q2 2015 were obtained from 42 Joint Commission certified comprehensive stroke centers and included 65,389 inpatient records relevant to measures CSTK-1, CSTK-2, CSTK-3 and CSTK-6; only centers reporting at least 6 months of data and with ≥ 30 denominator cases were included in this dataset. It is unclear how many of the participating hospitals used electronic records versus paper records.
  - Developers conducted several construct validation analyses, first hypothesizing a relationship between this measure and three other TJC stroke measures (specifically, that hospitals doing well on this measure also do well on the other measures) and then examining the degree of association between the measure results using the Pearson Correlation Coefficient. This is an appropriate method of score-level validation.
    - ***Hypothesis 1:*** Results from this measure should be positively correlated with results from the TJC CSTK 01 measure (National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients). This hypothesis was tested using the pilot testing data and using the Q1-2 2015 data.
    - ***Hypothesis 2:*** Results from this measure should be positively correlated with results from the TJC CSTK-03 measure (Severity Measurement Performed for Subarachnoid Hemorrhage (SAH) and Intracerebral Hemorrhage (ICH) Patients (Overall Rate)). This hypothesis was tested using the Q1-2 2015 data.
    - ***Hypothesis 3:*** Results from this measure should be positively correlated with results from the TJC CSTK-06 measure (Nimodipine Treatment Administered). This hypothesis was tested using the Q1-2 2015 data.
- Face validity of data elements was reported but did not address face validity of the measure score as a

representation of quality, as required for the criterion.

- Data element validity was assessed for accuracy and clarity by hospitals, but the data element validity criterion requires comparison to a gold standard.

Validity testing results:

- **Hypothesis *1: Pilot testing data:***

Pearson Correlation Coefficients, N = 66 Prob > |r| under H0: Rho=0

	CSTK-01	CSTK-02
Correlation Coefficient	0.25543	1.00000
P value	0.0385	

  - These results indicate a statistically significant positive correlation between the two measures, thus confirming the developers' hypothesis.

- ***Q1-2 2015 data: Hypotheses 1,2, and 3:*** Results for these analyses are presented in a correlation table. The correlation values are positive and statistically significant at the alpha=0.10 level, and therefore confirm the developers' hypotheses. These results indicate that this measure is most highly correlated with the CSTK 01 National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients measure.

Questions for the Committee:

- *Are the test samples adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

Patients are excluded from the measure if they are under age 18, have a length of stay greater than 120 days, were admitted for elective carotid intervention, or died during the inpatient stay.

The developers provide frequencies of exclusions for the pilot test data and the Q1-Q2 2015 data.

	Pilot test data	Q1-Q2 2015 data
<18	0.37%	--
LOS >120 days	0.11%	--
Admitted for Elective Carotid Intervention	0.35%	2.0%
Died during the stay	6.68%	4.0%

-- Q1-Q2 2015 dataset already excluded patients <18 or LOS>120 days

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- See data described under Performance Gap, above.
- The difference between the 10th and 25th percentiles indicates some variation between providers.

Question for the Committee:

o *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- Only one data source (the medical record) is specified.

2b7. Missing Data

- The developer describes the handling of missing data in the measure specifications (S.22) but does not provide information on the frequency of missing data or potential impact on measure results.

Guidance from algorithm: Specifications consistent with evidence (Box 1) → potential threats to validity mostly assessed (Box2) →empirical validity testing (Box 3) → measure score testing (Box 6) → sound conceptualization and hypotheses (Box 7) →high confidence that score are a valid indicator of quality(Box 8b) → Moderate

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2. Scientific Acceptability of Measure Properties

2a1. & 2b1. Specifications

Comments: **Evidence was not provided.
**The measure allows for a response of "8" = "mRS not performed" so the numerator could be maximized without actually recording patient data.

2a2. Reliability Testing

Comments: **The validity testing adequately linked the mRS instrument to other commonly used metrics however never addressed the absence of any linkage to outcomes.
**This is a bit conceptually strange given point #1. Even if a correlation exists with other quality measures, I don't know how one could argue that the mRS, per se, was the key component of quality as opposed to say a reflection of a latent construct of quality at the facility level
**The reliability testing was accurate

2b2. Validity Testing

Comments: **1) Not assessed is accepted so there will likely be site-to-site variation in whether subjects "not assessed" are submitted or not.
2) The absence of evidence to support the data collection makes validity difficult to assess.
**Exclusions are fairly reasonable. No risk adjustment exists. Without a good answer on evidence, I don't know what a meaningful difference would be. Missing data seems highly likely to be a major real world problem and one that is hard to address. Some hospitals are likely (no matter how hard they try) to have different rates of missingness based on their patient populations.
There probably should be risk adjustment here as well, but there is little data on how to do it. Some factors (i.e. severity, discharge location) likely influence how difficult it is to obtain a 90 day mRS.
**The validity testing adequately linked the mRS instrument to other commonly used metrics however never addressed the absence of any linkage to outcomes.

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **The reliability testing was adequate.
**Percent agreement for mRS is actually probably not a terrible reliability measure, because there is a lot of variation and thus chance variation is unlikely to be a huge problem.
I believe there is additional data that mRS (even by phone) can be reasonably reliably obtained.
**Not assessed is accepted so there will likely be site-to-site variation in whether subjects "not assessed" are submitted or not.
**The absence of evidence to support the data collection makes validity difficult to assess.

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- Data source is medical record abstraction: not all hospitals are able to generate the data electronically so a paper-based option is available
- Some data elements are in electronic form
- The measure is in use in the Joint Commission's Stroke Center Certification program.
- Data collection burden is significant:
  - Hospital staff resources to conduct the 90-day mRS
  - Medical record review to abstract all data elements required for the measure set (not just this measure) averaged 45 minutes per record.
  - Developer estimated "that the cost per case to abstract for this measure was approximately $3.50."

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Is the burden of data collection reasonable for a national performance measure? What resources are required to conduct the 90-day mRS?*
- *Are the required data elements available in electronic form (e.g., EHR or other electronic sources)?*
- *Is the data collection strategy ready to be put into operational use? Is conducting the mRS required for certification?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3. Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **The measure supposes that mRS is not routinely collected in the follow-up interval and so by definition this is a measure of data not routinely collected at some sites of care. The question is whether this would be reimbursable or not. If not, feasibility is questionable.
**This is a lot of burden for very little demonstrable (or conceptual) benefit. It is plausible that you could collect these data, in the subset of patients that follow-up within one's health system, via EHR, but for the broader population this is highly unlikely.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**
OR
Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- The measure currently is used for internal quality improvement for a TJC Disease-Specific Care Certification for Comprehensive Stroke Centers program, which includes 95 hospitals nationwide.
- Additional planned uses include public reporting and external benchmarking to other organizations, although no timeframe was noted.

Improvement results

- See section 1b above. Although this is a new measure, the developers note improvement from initial pilot testing results obtained in 2013.

Unexpected findings (positive or negative) during implementation: None reported

Potential harms: None reported.

Feedback:

- Because this measure only assesses whether or not the mRS assessment has been done, providers may score high on this measure even if their patients have poor outcomes, per the mRS.
- In late 2014, the Measure Applications Partnership (MAP) considered an ***outcome measure***—*Clinical Outcome post Endovascular Stroke Treatment: Patients with 90 day mRs score of 0 to 2 post endovascular stroke intervention*—for potential inclusion in the Medicare Shared Savings Program and the various clinician programs (PQRS/Physician Compare/Physician Feedback/Value-Based Payment Modifier). The MAP encouraged continued development of this measure. The MAP's rationale for this recommendation for the clinician programs was as follows:

  MAP acknowledged this rapidly changing and evolving practice and supports development of this outcome measure. Large programs are performing 1-2 cases/week. The measure may need clinical selection criteria or risk adjustment. MAP strongly recommends a similar outcome measure for all stroke patients.

  In general, the MAP was critical of assessment measures when outcome measures are needed.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4. Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **This measure is not usable in the context of routine care. Different sites may pursue missing subjects (versus entering 8) with different levels of diligence, which would bias a future study of the outcome impact of this measure.

Criterion 5: Related and Competing Measures

Related or competing measures

Related measures:

- CSTK-01 National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients;
- CSTK-03 Severity Measurement Performed for Subarachnoid Hemorrhage (SAH) and Intracerebral Hemorrhage (ICH) Patients (Overall Rate).
- CSTK-06: Nimodipine Treatment Administered

Harmonization

- All related measures are from the same developer and are harmonized**.**

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Title: CSTK-02: Modified Rankin Score (mRS) at 90 Days
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure title

Date of Submission: 1/15/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

Subcriterion 1a. Evidence to Support the Measure Focus

The measure focus is a health outcome or is evidence-based, demonstrated as follows:

- Health outcome:[3] a rationale supports the relationship of the health outcome to processes or structures of care.
- Intermediate clinical outcome, Process,[4] or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence[5] that the measure focus leads to a desired health outcome.
- Patient experience with care: evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information OR that patient experience with care is

correlated with desired outcomes.
- Efficiency:[6] evidence for the quality component as noted above.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement.

5. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

6. Measures of efficiency combine the concepts of resource use and quality (NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of:

Outcome

☐ Health outcome: Click here to name the health outcome

Health outcome includes patient-reported outcomes (PRO, i.e., HRQoL/functional status, symptom/burden, experience with care, health-related behaviors)

☐ Intermediate clinical outcome: Click here to name the intermediate outcome

☐ Process: Ishchemic stroke patients treated with intra-venous (IV) or intra-arterial (IA) thrombolytic (t-PA) therapy or who undergo mechanical reperfusion therapy for whom a 90 day (> 75 days and < 105 days) mRS is obtained via telephone or in-person

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME PERFORMANCE MEASURE *If not a health outcome, skip to 1a.3*

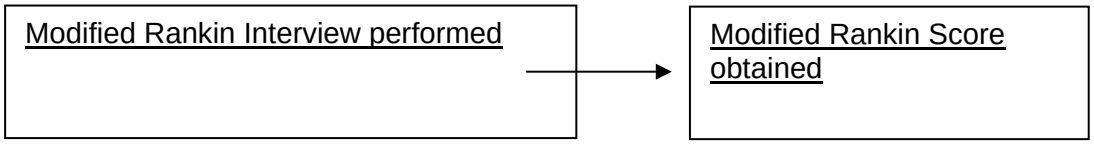
1a.2. Briefly state or diagram the linkage between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

Note: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the linkages between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



Standardized interviews conducted with the patient and/or caregiver yield a modified Rankin Score which is used to assess functional outcome and the extent of disability after stroke.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***
☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***
☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***
☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):
Schwamm LH, Holloway RG, Amarenco P, Audebert HJ, Bakas T, Chumbler NR, Handschu R, Jauch EC, Knight WA IV, Levine SR, Mayberg M, Meyer BC, Meyers PM, Skalabrin E, Wechsler LR. A review of the evidence for the use of telemedicine within stroke systems of care: A Scientific Statement from the American Heart Association/American Stroke Association. *Stroke.* 2009;40:2616-2634.

URL:
http://stroke.ahajournals.org/content/40/7/2635.full.pdf+html

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Rehabilitation: Feasibility and Effectiveness of Telephonic Consultation for Performing Assessments of Disability After Stroke (Page 2629)
Class I
1. Telephonic assessment for measuring functional disability after stroke is recommended when in-person assessment is impractical, the standardized rating instruments have been validated for telephonic use, and administration is by trained personnel using a structured interview.
(*Level of Evidence: B*)

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Class I: Procedure/treatment SHOULD be performed/administered.

Level of Evidence B: Limited populations evaluated. Data derived from a single randomized trial or nonrandomized studies.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system.
(*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

ACCF/AHA Classification of Recommendation and Level of Evidence
Classification Types
Class II a: It is reasonable to perform procedure/administer treatment.
Class II b: Procedure/Treatment may be considered.
Class III: Procedure/Treatment not helpful/no proven benefit/may be harmful.

Level of Evidence
Level A: Data derived from multiple randomized clinical trials or meta-analyses.
Level C: Only consensus opinion of experts, case studies, or standard of care.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Same as 1a.4.1

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Quinn TJ, Dawson J, Walters MW, Lees KR. Reliability of the Modified Rankin Scale: A Systematic Review. *Stroke.* 2009;40:3393-3395.

http://stroke.ahajournals.org/content/40/10/3393.full.pdf

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**
The systematic review identified quality of evidence based on bias and trial quality. Throughout the process, the reviewers adhered to Meta-analysis of Observational Studies in Epidemiology (MOOSE) guidelines for meta-analysis.

Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB for the Meta-analysis of Observational Studies in Epidemiology (MOOSE) Group. Meta-analysis of Observational Studies in Epidemiology: a proposal for reporting. *JAMA.* 2000;283:2008-2012.
http://jama.jamanetwork.com/article.aspx?articleid=192614

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?
The information in the following questions in this section is based on the studies included in the systematic review in section 1a.6.1.

The systematic review addressed the reliability of the modified Rankin Scale. To allow for comparison and where data permitted, the reviewers described results using *K,* quadratic weighted *K,* and percentage agreement. Based on previous work suggesting a beneficial effect of a structured interview approach, a separate analysis comparing "structured" and "traditional" mRS was performed. Interestingly, studies with larger numbers of patients and observers reported poorer reliability.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

An overall grade for the quality of quoted evidence was not assigned.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

An overall grade for the quality of quoted evidence was not assigned.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range
1988 - 2009

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

10 observational studies.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Overall quality of evidence across studies appears to be moderate. There were several limitations noted in the study design. First, there was heterogeneity between studies in several aspects of methodology. Collation of studies with differences in methodology potentially weakens the meta-analysis. Also the quality of studies varied for all those included, certain data were incomplete. Two studies measured mRS at distinct periods in the patient's recovery and were therefore subject to recall bias and the potential for functional improvement between assessments. Finally, no universally accepted analysis method for multiple *K* statistics has been described. Recognizing this limitation, the reviewers used a group analysis technique that made the fewest assumptions of the underlying data.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

This review demonstrates that overall reliability of mRS is moderate for the two approaches to mRS (i.e., traditional and structured). Inter-observer variability of mRS varied from "near perfect" (weighted *K*=0.95) to "poor" (*K*=0.25). Structuring mRS may partly improve mRS reliability, but the effects were not consistent.

Total 'n' = 587 patients
Median number of included patients = 47
Median number of researchers = 2

Reliability of Traditional mRS

Study	*K**	Weighted *K**	Agreement, %+
van Swieten, 1988	0.56 (0.45-0.68)	0.91 (0.71-1.00)	65%
Wolfe, 1991	N/A	0.87 (0.84-0.97)	80%
Berger, 1999	0.56 (0.41-0.71)	0.88 (0.58-1.00)	N/A
Wilson, 2005	0.44 (0.29-0.62)	0.78 (0.53-1.00)	57%
Newcommon, 2003	0.72 (0.55-0.89)	N/A	N/A
Wilson, 2005	0.25 (0.16-0.35)	0.71 (0.53-0.88)	43%
de Caneda, 2006	0.45 (0.31-0.60)	0.45 (0.58-0.90)	N/A
Gur, 2007	N/A	0.95 (0.89-1.00)	N/A
Meyer, 2008	N/A	0.90 (0.59-1.00)	N/A
Quinn, 2009	0.64 (0.48-0.79)	0.91 (0.65-1.00)	72%
Totals	0.46 (0.41-0.51)	0.90 (0.86-0.94)	71%

*95% CIs in parentheses
+Agreement between observers
N/A indicates not available

Reliability of mRS Using a Structured Interview Approach

Study	*K**	Weighted *K**	Agreement, %+
Wilson, 2002	0.70 (0.56-0.85)	0.93 (0.67-1.00)	78%
Newcommon, 2003	0.34 (0.17-0.55)	N/A	50%
Wilson, 2005	0.74 (0.64-0.84	0.91 (0.73-1.00)	81%
Quinn, 2009	0.50 (0.34-0.68)	0.74 (0.45-1.00)	63%
Totals	0.62 (0.56-0.69)	0.87 (0.75-1.00)	73%

*95% CIs in parentheses
+Agreement between observers
N/A indicates not available

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?
The evidence supports the reliability and feasibility of the modified Rankin Scale administered telephonically. Inter-observer variability is an inherent harm. Studies closest in design to large clinical trials demonstrated the greatest potential for variability. Inter-rater reliability may be improved by the use of experienced raters and structured interviews. Variability can also be controlled by limiting the use of a patient proxy.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable

1a.8 OTHER SOURCE OF EVIDENCE
If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

1a.8.2. Provide the citation and summary for each piece of evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
CSTK-02_Measure__Evidence_6.5.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
The Modified Rankin Scale (mRS) is the accepted standard for assessing recovery post-stroke. As such, it has become the most widely used clinical outcome measure for stroke clinical trials. Scores are used to measure the degree of disability or dependence in activities of daily living. Score reliability and reproducibility are improved through use of a structured interview by a trained evaluator. Interviews may be conducted in-person or over the phone (Schwamm, 2009). According to guideline recommendations from the American Heart Association/American Stroke Association, standardized interviews to obtain a mRS score should be conducted for acute ischemic stroke patients treated with IV or IA thrombolytic (t-PA) therapy or mechanical endovascular reperfusion therapy at 3 months (90 days) (Leifer, 2011); however, recovery may continue well beyond 3 months for many ischemic stroke patients.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Pilot Test Findings:
During the six-month pilot test (October 1, 2012-March 31, 2013), sixty-six sites submitted data for 10,218 completed patient records. For this measure, 1307 cases were assigned an ICD-9-CM Principal Diagnosis Code for ischemic stroke at Discharge and a procedure code associated with IV or IA thrombolysis or mechanical endovascular reperfusion therapy during the hospital stay. Of these denominator cases, 353 had an interview conducted at 90 days post-discharge to obtain a mRS score. The performance rates varied widely across sites for this measure with results ranging from a low of 0% to a high of 100%. The average rate for all sites collecting data for this measure was 27%, indicating a potential performance gap of approximately 73% if the optimal rate is 100%.

In January, 2015, The Joint Commission implemented data collection for the comprehensive stroke (CSTK) measure set to meet performance measurement requirements for its Comprehensive Stroke Certification Program. Below is the specified level of analysis for CSTK-02 Modified Rankin Score (mRS) at 90 Days for the two quarters of data received to date for this measure.

1Q 2015: 17 denominator cases; 10 numerator cases; 9 hospitals; 0.58824 national aggregate rate; 0.64815 mean of hospital rates; 0.44444 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 1.0 50th percentile rate/median rate; 0.33333 25th percentile rate/lower quartile; and, zero 10th percentile rate.

2Q 2015: 579 denominator cases; 506 numerator cases; 40 hospitals; 0.87392 national aggregate rate; 0.87638 mean of hospital rates; 0.19345 standard deviation; 1.0 90th percentile rate; 1.0 75th percentile rate/upper quartile; 0.97619 50th percentile rate/median rate; 0.8175 25th percentile rate/lower quartile; and, 0.57927 10th percentile rate.

1b.3. **If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.**
Citation:
Graham GD. Tissue Plasminogen Activator for Acute Ischemic Stroke in Clinical Practice: A Meta-Analysis of Safety Data. Stroke. 2003;34:2847-2850.

Summary:

Data analyzed: 1995-2003; 15 published, open-label studies (10 prospective, 5 retrospective or mixed); 2639 total patients. This review found that the reporting of outcomes was not standardized across studies. Individual studies used modified Rankin scores (mRS) of 0 to1 or 0 to 2, National Institutes of Health Stroke Scale (NIHSS) scores of 0 to 1, or Barthel Index scores of 95 to 100 to define the very favorable outcome group, whereas some studies provided no outcome data. Failure to report outcomes at a consistent time point (e.g., 30 or 90 days after stroke) or to use consistent criteria limited comparative assessment of treatment efficacy.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
This is the initial submission of this measure.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.
Kapral MK, Fang J, Hill MD, Silver F, Richards J, Jaigobin C, Cheung AM. Sex differences in stroke care and outcomes: results from the Registry of the Canadian Stroke Network. Stroke. 2005;36:809-814.

The study utilized data from the Registry of the Canadian Stroke Network to compare in-hospital and 6-month stroke outcomes in men and women. The study sample included 3323 patients (1527 female) with stroke hospitalized for acute care from July 2001 to February 2002 (Phase 1) and June to December 2002 (Phase 2). Although the median Rankin score at discharge was similar in women and men, there was a trend toward a higher proportion of women with severe disability (mRS score of 4 or 5) at discharge (30% versus 26%; P=0.0911). The study also found that women with stroke were slightly older than men, had a slightly longer length of stay, were more likely to be discharged to long-term care, and had slightly worse functional outcome at six-months after stroke. The median Stroke Impact Scale-16 (SIS-16) score at six months was 85.9 for women versus 92.2 for men; P=0.0001; score 100 best. A difference of 10 to 15 points on the SIS-16 is considered clinically significant and correlates with a difference of 1 point in the modified Rankin Scale.

_____

Petrea RE, Beiser AS, Seshadri S, Kelly-Hayes M, Kase CS, Wolf PA. Gender differences in stroke incidence and poststroke disability in the Framingham Heart Study. Stroke. 2009;40-1032-1037.

The study observed 1136 incident strokes (638 in women) over 56 years of follow-up. Women were significantly (P<0.01) more disabled before stroke and in the acute phase of stroke in dressing (59% versus 37%), grooming (57% versus 34%), and transfer from bed to chair (59% versus 35%). At 3 to 6 months post-stroke women were more disabled, more likely to be single, and 3.5 times more likely to be institutionalized (P<0.01).

_____

Persky RW, Turtzo LC, McCullough LD. National Institutes of Health (NIH) Author Manuscript. Stroke in women: disparities and outcomes. Curr Cardiol Rep. 2010;12(1):6-13.

The manuscript noted striking disparities between men and women respecting differences in disability and recovery post-stroke. More women than men were unable to independently perform various functional activities at 3-6 months post-stroke: eating 15% versus 9%; dressing 37% versus 20%; grooming 32% versus 17%; transfer from bed to chair 32% versus 13%; walking 32% versus 18% (Framingham Heart Study, 2009).

_____

Stansbury JP, Jia H, Williams LS, Vogel WB, Duncan PW. Ethnic disparities in stroke: epidemiology, acute care, and postacute outcomes. Stroke. 2005;36:374-386.

The review noted disparities in functional rehabilitation outcomes. In a prospective study of 145 patients (41 black stroke patients) drawn from a VA medical center, a university hospital, and a community sample in North Carolina black patients had greater physical impairments on admission that continued to be significantly worse at follow-up through 90 days with the greatest disparity at 30 days (P=0.002).

_____

Putman K, Horn S, Smout R, Dejong G, Deutscher D, Tian W, Hsieh CH. Racial disparities in stroke functional outcomes upon discharge from inpatient rehabilitation facilities. Disabil Rehabil. 2010;32(19):1604-11.

The multi-center prospective trial involved 732 patients from six inpatient rehabilitation facilities (IRFs) across the United States. In the moderate stroke group (N=397), functional scores (Functional Independence Measurement (FIM)) on admission were not significantly different between blacks and whites. In the severe stroke group (N=335), whites showed significantly lower functional scores at admission (mean score 44 versus 49 for blacks, p<0.001). Multivariate analyses predicting discharge FIM score found no significant differences between black and white stroke groups (p=0.2194 moderate stroke group; p=0.3547 severe stroke group.) The investigators concluded that the absence of significant differences in recovery while patients were on the rehabilitation unit suggests that racial disparities in long-term functional recovery after stroke are likely to have originated before or after the inpatient rehabilitation stay.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Each year, ~ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. Approximately 55,000 more women than men have a stroke. Of all strokes, 87% are ischemic strokes.

Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. According to 2010 mortality data, one of every 19 deaths in the United States is attributable to stroke; 60% of stroke deaths are women.

Stroke is also a leading cause of serious long-term disability in the United States (Survey of Income and Program Participation, a survey of the US Census Bureau). In 2010, stroke was among the top 18 diseases contributing to years lived with disability; of these 18 causes, only the age-adjusted rates for stroke increased significantly between 1990 and 2010 (P<0.05). Data from the National Heart, Lung and Blood Institute (NHLBI) revealed that 50% of ischemic stroke survivors age > 65 years had some hemiparesis; 46% had cognitive deficits, 35% experienced depressive symptoms; 30% were unable to ambulate without assistance; 26% were dependent in activities of daily living; 19% had aphasia; and, 26% were institutionalized in a nursing home. Among Medicare patients discharged from the hospital after stroke, approximately 45% are discharged to home (32% returning to home use home healthcare services), 24% to an inpatient rehabilitation facility, and 31% discharged to a skilled nursing facility. The mean lifetime cost of ischemic stroke, including inpatient care, rehabilitation, and follow-up as necessary for residual deficits are estimated at $140,048 per person.

The modified Rankin Scale (mRS) is a functional outcome scale used to assess the degree of disability or dependence in the activities of daily living post-stroke. Originally introduced by Rankin in 1957 and modified in 1988, the mRS at three months has since become the gold standard for measuring functional outcome after ischemic stroke.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

Leifer D, Bravata DM, Connors JJ III, Hinchey JA, Jauch EC, Johnston SC, Latchaw R, Likosky W, Ogilvy C, Qureshi AI, Summers D, Sung GY, Williams LS, Zorowitz R, on behalf of the American Heart Association Special Writing Group of the Stroke Council, Atherosclerotic Peripheral Vascular Disease Working Group and Council on Cardiovascular Surgery and Anesthesia, and Council on Cardiovascular Nursing. Metrics for measuring quality of care in comprehensive stroke centers: detailed follow-up to Brain Attack Coalition comprehensive stroke center recommendations: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2011;42:859.

Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Functional Status, Health and Functional Status : Functional Status

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.jointcommission.org/assets/1/6/CSTKManual2015August.pdf

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Copy_of_AppendixACSTKTables_ICD10codes-635878788267211970.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not Applicable

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Ischemic stroke patients for whom a 90 day (greater than or equal to 75 days and less than or equal to 105 days) mRS is obtained via telephone or in-person.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Two data elements are used to calculate the numerator. Data elements and definitions:
• Modified Rankin Score (mRS) – Documentation in the medical record of a mRS. The mRS is a 6 point disability scale with possible scores ranging from 0 to 5. A separate category of 6 is added for patients who expire. The mRS is the most widely used outcome measure in stroke clinical trials. Standardized interviews to obtain a mRS score are recommended at 3 months (90 days) following hospital discharge. Allowable Values: 0 (The patient has no residual symptoms); 1 (The patient has no significant disability; able to carry out all pre-stroke activities); 2 (The patient has slight disability; unable to carry out all pre-stroke activities but able to look

after self without daily help); 3 (The patient has moderate disability; requiring some external help but able to walk without the assistance of another individual); 4 (The patient has moderately severe disability; unable to walk or attend to bodily functions without assistance of another individual); 5 (The patient has severe disability; bedridden, incontinent, requires continuous care); 6 (The patient has expired); 7 (Unable to contact patient/caregiver); 8 (Modified Rankin Score not performed, OR unable to determine (UTD) from medical record documentation).

• Modified Rankin Score (mRS) Date – The month, day, and year that the mRS was obtained post-discharge.

Patients are eligible for the numerator population when the Modified Rankin Score (mRS) Date minus the Discharge Date is greater than or equal to 75 days and less than or equal to 105 days.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Ischemic stroke patients treated with IV or IA thrombolytic (t-PA) therapy or who undergo mechanical endovascular reperfusion therapy.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Included Populations:
• Discharges with ICD-10-CM Principal Diagnosis Code for ischemic stroke as defined in Appendix A, Table 8.1, AND
• Patients with documented thrombolytic (IV or IA t-PA) therapy (ICD-10-PCS Principal or Other Procedure Codes as defined in Appendix A, Table 8.1a), OR
• Patients with documented mechanical endovascular reperfusion therapy (ICD-10-PCS Principal or Other Procedure Codes as defined in Appendix A, Table 8.1b).

10 data elements are used to calculate the denominator. Data elements and definitions:

• Admission Date: The month, day, and year of admission to acute inpatient care.

• Birthdate: The month, day, and year the patient was born.

• Discharge Date - The month day and year the patient was discharged from acute care, left against medical advice or expired during the stay.

• Discharge Disposition – The final place or setting to which the patient was discharged on the day of discharge.

• Elective Carotid Intervention - Documentation demonstrates that the current admission is solely for the performance of an elective carotid intervention (e.g., elective carotid endarterectomy, angioplasty,carotid stenting). Allowable Values: Yes or No/UTD.

• ICD-10-PCS Other Procedure Codes: The International Classification of Diseases, Tenth Revision, Master Code Table (ICD-10-PCS) codes identifying all significant procedures other than the principal procedure.

• ICD-10-PCS Other Procedure Dates: The month, day, and year when the associated procedure(s) was (were) performed.

• ICD-10-CM Principal Diagnosis Code: The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.

• ICD-10-PCS Principal Procedure Code: The International Classification of Diseases, Tenth Revision, Master Code Table (ICD-10-PCS) code that identifies the principal procedure performed during this hospitalization. The principal procedure is the procedure performed for definitive treatment rather than diagnostic or exploratory purposes, or which is necessary to take care of a complication.

• ICD-10-PCS Principal Procedure Date: The month, day, and year when the principal procedure was performed.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
• Patients less than 18 years of age
• Patients who have a Length of Stay greater than 120 days
• Patients admitted for Elective Carotid Intervention
• Patients who expire during the hospital stay

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
• Patients less than 18 years of age.
o Patient age (in years) equals Admission Date minus Birthdate.

• Patients who have a Length of Stay greater than 120 days.
o Length of Stay (in days) equals Discharge Date minus Admission Date.

• Patients admitted for Elective Carotid Intervention:
o Elective Carotid Intervention includes procedures of the head and neck as defined in Appendix A, Table 8.3 Carotid Intervention Procedures when medical record documentation also states that the reason for the patient's admission to the hospital was for the performance of that procedure and not for the treatment of acute ischemic stroke.
o An elective admission is documented as a pre-planned or scheduled admission to the hospital.

• Patients who expire during the hospital stay
o Determined by the data element Discharge Disposition, allowable value #6 Expired

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Not Applicable

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not Applicable

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
Not Applicable

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including*

identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)
Comprehensive Stroke (CSTK) Initial Patient Population Algorithm

Variable Key: Patient Age, Initial Patient Population Reject Case Flag, Length of Stay, Sub-Population 1 Flag, Sub-Population 2 Flag, and Sub-Population 3 Flag.

1. Start CSTK Initial Patient Population logic sub-routine. Process all cases that have successfully reached the point in the Transmission Data Processing Flow: Clinical which calls this Initial Patient Population Algorithm. Do not process cases that have been rejected before this point in the Transmission Data Processing Flow: Clinical.

2. Check ICD-10-CM Principal Diagnosis Code
a. If the ICD-10-CM Principal Diagnosis Code is not on Table 8.1 and 8.2, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the ICD-10-CM Principal Diagnosis Code is on Table 8.1 or 8.2, continue processing and proceed to the Patient Age calculation.

3. Calculate Patient Age. Patient Age, in years, is equal to the Admission Date minus the Birthdate. Use the month and day portion of admission date and birthdate to yield the most accurate age.

4. Check Patient Age
a. If the Patient Age is less than 18 years, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the Patient Age is greater than or equal to 18 years, continue processing and proceed to Length of Stay Calculation.

5. Calculate the Length of Stay. Length of Stay, in days, is equal to the Discharge Date minus the Admission Date.

6. Check Length of Stay
a. If the Length of Stay is greater than 120 days, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the Length of Stay is less than or equal to 120 days, the patient is in the CSTK Initial Patient Population.

7. Set the Initial Patient Population Reject Case Flag to equal No. Continue processing and proceed to the ICD-10-CM Principal Diagnosis Code to determine the CSTK sub-population.

8. Initialize Sub-Population 1 Flag, Sub-Population 2 Flag and Sub-Population 3 Flag to No.

9. Check ICD-10-CM Principal Diagnosis Code
a. If the ICD-10-CM Principal Diagnosis Code is on 8.2, the patient is in the CSTK Sub-Population 3 and is eligible to be sampled for the CSTK Sub-Population 3. Set Sub-Population 3 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If the ICD-10-CM Principal Diagnosis Code is on 8.1, continue processing and proceed to ICD-10-PCS Principal or Other Procedure Codes.
i. If at least one ICD-10-PCS Principal or Other Procedure Codes is on Table 8.1a or 8.1b, the patient is in the CSTK Sub-Population 2 and is eligible to be sampled for the CSTK Sub-Population 2. Set Sub-Population 2 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
ii. If none of the ICD-10-PCS Principal Or Other Procedure Codes are on Table 8.1a or 8.1b, the patient is in the CSTK Sub-Population 1 and is eligible to be sampled for the CSTK Sub-Population 1. Set Sub-Population 1 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

CSTK-02: Modified Rankin Score (mRS) at 90 Days

Numerator: Ischemic stroke patients for whom a 90 day (> 75 days and < 105 days) mRS is obtained via telephone or in-person.

Denominator: Ischemic stroke patients treated with IV or IA thrombolytic (t-PA) therapy or who undergo mechanical reperfusion

therapy

Variable Key: Days
1. Start processing. Run cases that are included in the Comprehensive Stroke (CSTK) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check ICD-10-CM Principal Diagnosis Code
a. If ICD-10-CM Principal Diagnosis Code is not on Table 8.1, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
b. If ICD-10-CM Principal Diagnosis Code is on Table 8.1, continue processing and proceed to Elective Carotid Intervention.

3. Check Elective Carotid Intervention
a. If Elective Carotid Intervention is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Elective Carotid Intervention equals Yes, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If Elective Carotid Intervention equals No, continue processing and proceed to Discharge Disposition.

4. Check Discharge Disposition
a. If Discharge Disposition equals 6, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
b. If Discharge Disposition equals 1, 2, 3, 4, 5, 7, or 8, continue processing and proceed to ICD-10-PCS Principal or Other Procedure Codes.

5. Check ICD-10-PCS Principal or Other Procedure Codes
a. If all ICD-10-PCS Principal or Other Procedure Codes is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing
b. If none ICD-10-PCS Principal or Other Procedure Codes is on Table 8.1a or 8.1b, the case will proceed to a Measure Category Assignment of B and will not be in the measure population. Stop processing.
c. If any ICD-10-PCS Principal or Other Procedure Codes is on Table 8.1a or 8.1b, continue processing and proceed to Modified Rankin Score (mRS).

6. Check Modified Rankin Score (mRS)
a. If Modified Rankin Score (mRS) is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Modified Rankin Score (mRS) equals 8, the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
c. If Modified Rankin Score (mRS) equals 0, 1, 2, 3, 4, 5, 6, or 7, continue processing and proceed to Modified Rankin Score (mRS) Date.

7. Check Modified Rankin Score (mRS) Date
a. If Modified Rankin Score (mRS) Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Modified Rankin Score (mRS) Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to Discharge Date.

8. Check Discharge Date
a. If Discharge Date is missing, the case will proceed to a Measure Category Assignment of X and will be rejected. Stop processing.
b. If Discharge Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to the Days Calculation.

9. Days Calculation. Days calculation, in days, is equal to the Modified Rankin Score (mRS) Date minus the Discharge Date.
a. If the number of days is less than 75 or greater than 105, the case will proceed to a Measure Category Assignment of D and will be in the measure population. Stop processing.
b. If the number of days is greater than or equal to 75 and less than or equal to 105, the case will proceed to a Measure Category Assignment of E and will be in the numerator population. Stop processing.

S.19. Calculation Algorithm/Measure Logic Diagram URL or Attachment *(You also may provide a diagram of the Calculation*

Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter for the measure set cannot sample. Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

The sub-population for the CSTK-02 measure Initial Patient Population is CSTK 2-Ischemic Stroke with IV t-PA, IA t-PA, or MER. The CSTK 2 sub-population sampling group includes patients admitted to the hospital for inpatient acute care if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.1 AND ICD-10-PCS Principal or Other Procedure Codes as defined in Appendix A, Table 8.1a OR Table 8.1b, a Patient Age (Admission Date – Birthdate) > 18 years and a Length of Stay (Discharge Date - Admission Date) less than or equal to 120 days.
Quarterly Sampling

The Quarterly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for CSTK sub-population 2. Hospitals performing quarterly sampling for the CSTK-02 measure must ensure that its Initial Patient Population and sample size meet the following conditions for the CSTK 2 sub-population sampling group:

Quarterly Sample Size Based on CSTK Sub-population 2 for Ischemic Stroke with IV t-PA, IA t-PA, or MER (Table 2)

Sub-Population 2: If "N" > 420, then 'n' 84
Minimum Required Sample Size: 84 records

Sub-Population 2: If "N" 211-419, then 'n' 20%
Minimum Required Sample Size: 20% of Sub-Population size records

Sub-Population 2: If "N" 43-210, then 'n' 42
Minimum Required Sample Size: 42 records

Sub-Population 2: If "N" < 42, then 'n' 100%
Minimum Required Sample Size: No sampling; 100% Sub-Population records required

Monthly Sampling
The Monthly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for CSTK sub-population 2. Hospitals performing monthly sampling for the CSTK-02 measure must ensure that its Initial Patient Population and sample size meet the following conditions for each sub-population sampling group:

Monthly Sample Size Based on CSTK Sub-population 2 for Ischemic Stroke with IV t-PA, IA t-PA, or MER (Table 5)

Sub-Population 2: If "N" > 140, then 'n' 28
Minimum Required Sample Size: 28 records

Sub-Population 2: If "N" 71-140, then 'n' 20%
Minimum Required Sample Size: 20% of Sub-Population records

Sub-Population 2: If "N" 15-70, then 'n' 14
Minimum Required Sample Size: 14 records

Sub-Population 2: If "N" < 14, then 'n' 100%
Minimum Required Sample Size: No sampling; 100% Sub-Population records required

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not Applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
A web-based data collection tool was developed by The Joint Commission for the pilot test process. Currently, hospitals have the flexibility of creating their own tool modeled after the pilot tool or they may develop their own data collection tools using the data element dictionary and allowable values specified in the implementation guide. Hospitals also have the option of selecting a vendor-developed data collection tool which has been verified by The Joint Commission.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not Applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2865_MeasureTesting_MSF6.5.docx

Measure Title: CSTK-02: Modified Rankin Score (mRS) at 90 Days
Date of Submission: 1/15/2016
Type of Measure: Process

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions.
15. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.
16. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Not Applicable

1.3. What are the dates of the data used in testing? October 1, 2012 – July 15, 2013; first and second quarter 2015.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Hospitals were recruited for the pilot test of the measures via an open call on The Joint Commission web site. An announcement of the call with a link to The Joint Commission web site was also posted on the American Heart Association/American Stroke Association web site. Hospitals were selected with the intent to capture variability related to ownership, size, type of facility and location. Eighty-two hospitals from twenty-seven states were selected from more than 120 volunteers to participate in the six-month pilot test of the measures. Twenty hospitals withdrew during the pilot test citing lack of resources to complete the project. Sixty-two hospitals submitted data for each month of the six-month pilot test. An additional four hospitals submitted data for one or more months.

Sixty-six hospitals contributed data for the analysis of the measures:

Ownership:

For Profit	14
Not for Profit	52

Bed Size:	
Less than 100	0
100 – 199	3
200 – 299	7
300 – 499	26
Greater than 500+	30

Located in 27 states:
Alabama
Arizona
California
Colorado
Florida
Georgia
Illinois
Indiana
Kentucky
Louisiana
Maryland
Massachusetts
Michigan
Minnesota
Missouri
Nevada
New Jersey
New York
North Carolina
Ohio
Oregon
Pennsylvania
Tennessee
Texas
West Virginia
Washington
Wisconsin

Other:	
Teaching	41
Non-teaching	25

Urban	59
Rural	7

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)?
(*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

During the six-month pilot test, sixty-six hospitals submitted data for 1,307 inpatient records. The cases included patients greater than 18 years of age, male and female, all races, and all payers. Age, gender, racial, and payer distribution are not known because the results were de-identified. The patient population was comprised of all ischemic stroke patients treated with intra-venous (IV) or intra-arterial (IA) thrombolytic (t-PA) therapy or mechanical endovascular reperfusion therapy.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Data from the six month pilot test of the measures were used for reliability testing and face validity. To test the empirical validity of the measures, two quarters of data from 42 Joint Commission certified comprehensive stroke centers was used to conduct a secondary analysis.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Reliability testing was performed at twelve participating pilot sites. Testing was conducted on a stratified random sample of records selected from each organization at the organization and measure category level. Hospitals were visited by teams of two Joint Commission staff during April, May, June, July, and August 2013. During these visits, Joint Commission staff re-abstracted records previously abstracted by hospital staff. A total of 281 records were re-abstracted. In cases of disagreement between the re-abstracted and original abstraction on a data element, reasons for disagreement were determined and adjudication was made as to whether original or re-abstraction findings were correct. Reliability was addressed by comparing the original abstracted and adjudicated re-abstracted values, with the adjudicated value serving as the gold standard. The data analysis included both the percent agreement and the kappa statistic to adjust for chance agreement for categorical data elements.

2a2.3. For each level checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data Element	Number of Mismatches	Match Rate	Kappa
Elective Carotid Intervention	3	98.9%	-0.0043
Modified Rankin Score (mRS)	42	85.1%	NA
Modified Rankin Score (mRS) Date	23	91.8%	NA

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

A statistical measure of inter-rater reliability is Cohen's Kappa, which ranges generally from 0.0 to 1.0 (although negative numbers are possible), where large values mean better reliability and values near zero suggest that agreement is attributable to chance alone. It indicates the proportion of agreement not expected by chance alone (e.g., Kappa of 0.6 means that raters agreed 60% of the time over and above what would be expected by chance alone).

Landis & Koch, 1977 offers the following clarification of Kappa interpretation:

< 0	Poor agreement
0.00-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-1.00	Almost perfect agreement

Other authors (Cicchetti & Sparrow; Fleiss) have suggested additional classifications for interpreting the Kappa statistic, but all seem to indicate Kappa > 0.60 is desirable.

The statistical measure of inter-rater reliability is suitable for the categorical data type, therefore the measure is applied to the data element of *Elective Carotid Intervention*. Although the Kappa value indicates poor agreement for *Elective Carotid Intervention*, the low marginal proportion of those with this intervention in the sample (<1%) means that Kappa in this case is not well determined. According to the statistical measure of Kappa value, we believe these results demonstrate acceptable reliability of the assessment data used in the performance measure.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

X **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

X **Empirical validity testing**

X **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Face Validity:

Measure face validity was assessed via survey and focus groups of hospitals participating in the pilot test. Focus group discussions were held at all test sites visited, during which we received feedback as to whether the measure, data elements, and definitions accurately reflected existing evidence. All of the respondents indicated that all aspects of the measures accurately reflected current evidence. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant.

To determine feasibility and identify areas for potential revision, test sites were asked to electronically rate the clarity of numerator statements, denominator statements, and measure information forms (MIFs) on a five point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree). Data elements and associated tables were evaluated for clarity, accuracy, data availability and accessibility.

Empirical Validity:

Measure convergent validity was assessed using hospitals patient level data from The Joint Commission warehouse. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous face validity testing. We conducted a secondary analysis to help interpret results; correlation among CSTK process measures.

The data were comprised of first and second quarter 2015 submissions. This included 42 hospitals submitting 65,389 inpatient records for the selected CSTK measures CSTK-1, CSTK-2, CSTK-3 and CSTK-6. The hospital's selection was based on those hospitals that reported 6 months of data and had 30 or more denominator cases for the period.

Comprehensive Stroke (CSTK) Initial Patient Population
The CSTK Initial Patient Population is unique in that it is comprised of three distinct subpopulations: ischemic stroke patients who do not undergo a reperfusion therapy (i.e., procedure), ischemic stroke patients who undergo a reperfusion therapy (IV t-PA, IA t-PA, or mechanical endovascular reperfusion (MER) therapy), and hemorrhagic stroke patients.

Subpopulation 1: Ischemic Stroke Without Procedure
This subpopulation comprises ischemic stroke patients who are admitted to the hospital for inpatient acute care and do not undergo a reperfusion procedure (CSTK-01 measure). Patients are included in the CSTK-1 Ischemic Stroke Without Procedure subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.1, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

Subpopulation 2: Ischemic Stroke With IV t-PA, IA t-PA, or MER
This subpopulation comprises ischemic stroke patients who receive IV t-PA, IA t-PA, or MER procedures during the hospital stay (CSTK-01 and CSTK-02 measures). Patients admitted to the hospital for inpatient acute care are included in the CSTK-2 Ischemic Stroke With IV t-PA, IA t-PA, or MER subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.1 and ICD-10-PCS Principal or Other Procedure Codes as defined in Appendix A, Table 8.1a OR Table 8.1b, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

Subpopulation 3: Hemorrhagic Stroke
This subpopulation comprises hemorrhagic stroke patients admitted to the hospital for inpatient acute care (CSTK-03 and CSTK-06 measures). Patients are included in the CSTK-3 -Hemorrhagic Stroke subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.2, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Pilot Test Data:
Descriptive statistics for the measure: N= 66 hospitals
Overall rate=23% (SD=30%), min=0%, max=100%

Additional analyses:
CSTK-01 and CSTK-02 are measures for ischemic stroke. The measure rate among these measures are expected to be positively correlated. The Pearson correlation coefficient was calculated.

Pearson Correlation Coefficient interpretation: the range is from -1 to 1. When the correlation coefficient is close to +1 or -1, it means that there is strong correlation; p value is utilized to determine if the corrlation coefficient is significant or not. If it is less than 0.05, then the correlation is considered significant.

The table below demonstrates that the CSTK-02 rate was positively correlated to the CSTK-01 measure.

Pearson Correlation Coefficients, N = 66 Prob > |r| under H0: Rho=0

	CSTK-01	**CSTK-02**
Correlation Coefficient	0.25543	1.00000
P value	0.0385	

The average rating for measure CSTK-02 numerator and denominator statements, including the clarity of numerator and denominator inclusions and exclusions was 4.26, indicating that the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

Data element validity was evaluated for clarity and accuracy, as well as, data availability and accessibility. The percentage of agreement for new data elements developed specifically for CSTK-02 and other measures in the CSTK measure set is detailed in the table below:

Data Element Name	Clarity / Accuracy	Availability / Accessibility
Modified Rankin Score (mRS)	97.67%	25.58%
Modified Rankin Score (mRS) Date	97.62%	25.58%

1Q and 2Q 2015 Data:
Overall descriptive statistics for CSTK selected measures: N = 42 certified comprehensive stroke hospitals: n = 65,389

CSTK-01
Median: 89%
Percentile 10%: 70%
Percentile 25%: 80%
Percentile 75%: 93%
Percentile 90%: 95%

CSTK-02
Median: 95%
Percentile 10%: 63%
Percentile 25%: 75%
Percentile 75%: 100%
Percentile 90%: 100%

CSTK-03
Median: 61%
Percentile 10%: 33%
Percentile 25%: 48%
Percentile 75%: 79%
Percentile 90%: 89%

CSTK-06
Median: 86%
Percentile 10%: 74%
Percentile 25%: 80%
Percentile 75%: 93%
Percentile 90%: 100%

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
CSTK_1	42	0.83687	0.17097	35.14857	0.00267	1.00000
CSTK_2	41	0.85894	0.20516	35.21648	0.20000	1.00000
CSTK_3	42	0.59308	0.22979	24.90933	0.07273	0.92969
CSTK_6	42	0.83555	0.17123	35.09314	0	1.00000

Pearson Correlation Coefficients Prob > |r| under H0: Rho=0 Number of Observations				
	CSTK_1	CSTK_2	CSTK_3	CSTK_6
CSTK_1	1.00000 42	0.41205 0.0074 41	0.65340 <.0001 42	0.69186 <.0001 42
CSTK_2	0.41205 0.0074 41	1.00000 41	0.28658 0.0693 41	0.31724 0.0433 41
CSTK_3	0.65340 <.0001 42	0.28658 0.0693 41	1.00000 42	0.55910 0.0001 42
CSTK_6	0.69186 <.0001 42	0.31724 0.0433 41	0.55910 0.0001 42	1.00000 42

The CSTK measures table shows a positive correlation and statistical significance which indicates that hospitals with high quality on one CSTK measure tend to have high correlations on the other stroke measures.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)
Pilot Test Data:
We note that ratings for both data elements and the measure are relatively high; numerators, denominators and measures were ranked above the midpoint of 3.0 (average), and data elements were above 75% positive in clarity, collectability, and correctness of data sources. We conclude that the measures and specifications are valid.

Since a process for follow-up at 90 days was not established at many sites until after initiation of data collection for the pilot test, 90 day mRS data were not always available. We expect that data availability will improve with use over time.

1Q and 2Q 2015 Data:
Overall the positive inter-correlations indicates convergent validity of all the measures.
They are positively correlated with other evidence-based processes of care.

2b3. EXCLUSIONS ANALYSIS .
NA ☐ no exclusions — *skip to section 2b4*
2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Pilot Test Data:
The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patients who have a Length of Stay greater than 120 days
- Patients admitted for Elective Carotid Intervention
- Patients who expire during the hospital stay

1Q and 2Q 2015 Data:
The following exclusions were analyzed by subpopulation and measure for frequency and variability across providers:

- Patients less than or equal to 18 years of age
- Patients who have a Length of Stay greater than or equal to 120 days
- Patients admitted for *Elective Carotid Intervention*
- Patients who expire during the hospital stay

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Pilot Test Data:
There were 10668 admissions included in the initial cohort. From among the 10668 admissions in 66 hospitals, the descriptive statistics are given below.

Exclusion: Patient not in CSTK Initial Patient Population
Note: A case was excluded from the Initial Patient Population as determined by the following:

- Patients less than 18 years of age: Overall Occurrence n = 39 (0.37%)
- Patients who have a Length of Stay greater than 120 days: Overall Occurrence n = 12 (0.11%)

There were 6508 admissions included in the initial cohort and diagnosed ischemic stroke. From among the 6508 admissions in 66 hospitals, the descriptive statistics are given below.

Exclusion: Elective Carotid Intervention
Overall Occurrence n = 23
Overall Occurrence Percentage: 0.35**%**

Exclusion: Patients who expire during the hospital stay
Overall Occurrence n= 435
Overall Occurrence Percentage: 6.68 **%**

1Q and 2Q 2015 Data:
There were 65,389 admissions selected from the initial cohort. From among the 65,389 admissions in 42 hospitals, the descriptive statistics are given below.

<u>Applied To Measure CSTK-01 CSTK-02</u>
Exclusion: Patients admitted for *Elective Carotid Intervention*
Overall Occurrence n = 602
Overall Occurrence Percentage: 2%
Minimum: 0.24%
Median: 3.0%
Maximum: 7.64%

Applied To Measure CSTK-02
Exclusion: Patients who expire during the hospital stay
Overall Occurrence n = 592
Overall Occurrence Percentage: 4%
Minimum: 1.34%
Median: 5%
Maximum: 7.84%

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Pilot Test Data:
According to the overall occurrences in 2b3.2, the overall frequency of exclusions is low for those in the measure denominator. The distribution of exclusions across hospitals is very narrow indicating that the occurrence is random and likely would not bias performance results.

It is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a Length of Stay greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Patients admitted for Elective Carotid Intervention
Rationale: Patients are excluded because the purpose of admission to the hospital was for the performance of an intervention to prevent stroke.

Patients who expire during the hospital stay
Rationale: It is not necessary to interview family/caregivers at 90 days for patients who expire during the hospital stay.

1Q and 2Q 2015 Data:
The overall frequency of exclusions is low for those in the measure denominator. The distribution of exclusions across hospitals is narrow indicating that the occurrence is random and likely would not bias performance results.

It is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a *Length of Stay* greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Patients admitted for *Elective Carotid Intervention*
Rationale: Patients are excluded because the purpose of admission to the hospital was for the performance of an intervention to prevent stroke.

Patients who expire during the hospital stay
Rationale: Patients who expire are not eligible to be in this measure

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

Not Applicable

2b4.1. What method of controlling for differences in case mix is used?
X **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☐ **Stratification by** Click here to enter number of categories **risk categories**
☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
Not Applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care and not related to disparities*)
Not Applicable

2b4.4. What were the statistical results of the analyses used to select risk factors?
Not Applicable

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)
Not Applicable

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
if stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
Not Applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Not Applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Not Applicable

2b4.9. Results of Risk Stratification Analysis:
Not Applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not Applicable

***2b4.11. Optional Additional Testing for Risk Adjustment** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*
Descriptive statistics for the performance measure scores for all tested entities were constructed. These statistics were the mean, standard deviation, median, minimum, and maximum scores.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)
A meaningful difference was defined as a spread of more than 30 percentage points between minimum and maximum scores or between median and maximum pilot hospital rates.

Results for CSTK-02 were:
Descriptive statistics for measure: N=66 hospitals
Overall rate=22.7%
Standard Deviation=30.3%
Minimum=0%
Median=0%
Maximum=100%

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

The results are interpreted as showing a meaningful spread between both the median and maximum scores and between minimum and maximum scores and indicates much room for improvement in performance.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Not Applicable

Note: *This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the*

numerator). ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different datasources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not Applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not Applicable

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i*.e., what do the results mean and what are the norms for the test conducted*)

Not Applicable

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other

If other: Data element allowable values are selected, either manually or electronically, from clinical and coded data available in medical record documentation or other sources, e.g., logs or spreadsheets that record follow-up data. All medical record documentation is used in the abstraction process. Vendor data collection tools are used to import data elements needed for measure rate calculation.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract this measure electronically, so offers a chart-abstracted version which allows for data capture from unstructured data fields. The Joint Commission plans to retool the measure for capture from electronic sources within the next several years.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

<u>IF a PRO-PM</u>, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

This measure was collected during a 6 month pilot process for Joint Commission Comprehensive Stroke Certification. It was learned via discussions and written pilot evaluation materials that this measure could be retained as written with minimal clarification of the following data element:

- Modified Rankin Score (mRS)
  - Allowable value '6' was selected for patients who expired during the hospital stay, although these patients are excluded from the denominator population.

This issue related to data abstraction has been easily resolved through clarification of guidelines for abstraction.

Other information impacting the feasibility and implementation of the measure was also obtained from the pilot process and is summarized as follows:

Staff Training and Education:
To prepare for and support continuous data collection throughout the pilot test, a total of ten hours was spent on staff training and education. Training was accomplished via two 2-hour webinars and monthly conference calls with pilot site participants. Some sites spent additional time training interviewers to conduct the standardized interviews needed to obtain a Modified Rankin Score.

Case Identification/Medical Record Retrieval:
Case identification was not a problem; cases were identified by the ICD-9-CM principal diagnosis codes for ischemic stroke. Record retrieval time varied depending on the type of medical record. On average, 10 minutes were spent for record retrieval with more time spent to retrieve a paper record than electronic health record.

Case Selection:
For the pilot test of the measure, sampling was allowed. The sampling methodology was modified post-pilot and sampling requirements established specifically for the sub-population of ischemic stroke patients who receive IV or IA thrombolytic therapy or mechanical reperfusion therapy.

Data Abstraction:
Medical record review to abstract all data elements required for the measure set averaged 45 minutes per record. Time spent on record review varied with case complexity and the number of procedures and interventions performed, as well as, the number of data elements collected for the measure. In general, ischemic stroke cases required more time for review than did hemorrhagic stroke cases.

Data abstraction was primarily done by nurses, (e.g., Registered Nurse(s) with a quality improvement background, Stroke Coordinators, and Advanced Practice Nurses). Some pilot sites reported that the abstractor reviewed the record with the medical director or neurologist at least initially to identify documentation of measure specific data elements. Data specialists or administrative staff were utilized to enter abstracted data into the on-line data collection tool.

Data Availability:
Data availability was an issue at the onset of the pilot test since many sites did not have a process in place to conduct the standardized interviews needed to obtain Modified Rankin Scores at 90 days. Data availability improved throughout the pilot test period as sites implemented internal processes to obtain these data either in-person or over the phone.

Cost of Data Abstraction:
Using 2012 national wage averages, it is estimated that the cost per case to abstract for this measure was approximately $3.50.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
This measure is one in a set of measures implemented with discharges effective January 1, 2015 for Joint Commission Comprehensive Stroke Certification. There are no fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Quality Improvement (Internal to the specific organization) Disease-Specific Care Certification for Comprehensive Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

• Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Not Applicable

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
Not Applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

Significant improvement has been noted since the pilot test of the measure based on the first two quarters of Joint Commission ORYX performance measurement data. The initial performance measure gap of approximately 73% has decreased more than 50% as demonstrated by a national aggregate rate of 87% (N=40) for 2Q2015; however, a greater than 40% gap exists for the lowest decile of hospitals.

4b.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of

initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
Not Applicable

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no unintended negative consequences reported or detected during testing or since implementation of the measure specifications.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Not applicable

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Karen, Kolbusz, kkolbusz@jointcommission.org, 630-792-5931-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Ann, Watt, awatt@jointcommission.org, 630-792-5944-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel was to provide advisory oversight in the literature review, measure construct and content, review of testing results, and endorsement of draft and finalized measures. Additionally they may be called upon in the future to provide measure content oversight and updates.

Technical Advisory Panel

Harold P. Adams, Jr., MD
Professor of Neurology
Director of the Division of Cerebrovascular Diseases
Director of the Stroke Center
University of Iowa Health Care
Iowa City, IA

Mark J. Alberts, MD
Clinical Vice-Chair
Department of Neurology and Neurotherapeutics
The University of Texas Southwestern Medical Center
Dallas, TX

Anne W. Alexandrov, PhD, RN, FAAN
Health Outcomes Institute
Fountain Hills, AZ

Andrew W. Asimos, MD
Director of Emergency Stroke Care

Carolinas Medical Center
Charlotte, NC

Thomas P. Bleck, MD, FCCM
Professor of Neurological Science, Neurosurgery, Internal Medicine, and Anesthesiology
Associate Chief Medical Officer, Critical Care
Rush Medical College
Chicago, IL

Darlene Bradley, RN, PhDc, MSN, MAOM
Director Emergency/Trauma Services
University of California Irvine Medical Center
Orange, CA

Mary G. George, MD, MSPH FACAS
Medical Officer
Division for Heart Disease and Stroke Prevention
Centers for Disease Control and Prevention
Atlanta, GA

Martin Gizzi, MD, PhD
Legacy Health System
Portland, OR

James, Grotta, MD
Professor and Chairman, Department of Neurology
Chief, Vascular Neurology
University of Texas Medical School-Houston
Houston, TX

Edward C. Jauch, MD, MS
Professor, Division of Emergency Medicine and Department of Neurosciences
Medical University of South Carolina
Charleston, SC

Irene Katzan, MD, MS
Director, Neurological Institute Center for Outcomes Research and Evaluation
Cleveland Clinic
Cleveland, OH

Jawad F. Kirmani, MD
Director of Stroke and Cerebrovascular Program
JKF Medical Center
Edison , NJ

Brajesh K. Lal, MD
University of Maryland Medical Center,
Division of Vascular Surgery
Baltimore, MD

Kimberly E. Levasseur-Franklin, PharmD, BCPS
Critical Care Pharmacist, Neuro/Spine ICU
Clinical Leader, ICU Pharmacist Work Group
Northwestern Memorial Hospital
Chicago, IL

Barbara Lutz, PhD, RN, CRRN, APHN-BC, FAHA, FNAP, FAAN
McNeill Distinguished Professor
School of Nursing / College of Health and Human Services
University of North Carolina-Wilmington
Wilmington, NC

Cameron McDougall, MD, FRCSC
Barrow Neurological Institute
Phoenix, AZ

Kathy Morrison, RN, MSN, CNRN
Stroke Program Coordinator
Penn State Hershey Medical Center
Hershey, PA

Charles J. Prestigiacomo, MD, FAANS, FACS
Professor, Department of Neurological Surgery, Radiology, Neurology and Neurosciences
Director, Residency Program in Neurological Surgery
Director, Cerebrovascular and Endovascular Surgery
University of Medicine and Dentistry of New Jersey
Newark, NJ

Howard A. Rowley, MD
Chief of Neuroradiology
Joseph Sackett Professor of Radiology
Professor of Radiology, Neurology, and Neurosurgery
University of Wisconsin
Madison, WI

Jeffrey Saver, MD, FAHA, FAAN
Professor of Neurology
Stroke Center Director
Professor of Medicine
David Geffen School of Medicine at UCLA,
Ronald Reagan UCLA Medical Center
Los Angeles, CA

Lee H. Schwamm, MD, FAHA
Vice Chair of Neurology, Professor of Neurology
Harvard Medical School
Chair, American Heart Association GWTG Steering Committee
Director, Stroke Program/Telemedicine

Massachusetts General Hospital
Boston, MA

Deborah, Summers, MSN, ACNS-BC, CNRN, CCRN, FAHA
Stroke Program Coordinator
St. Luke's Brain and Stroke Institute
Kansas City, MO

Dileep R. Yavagal, MD
Director, Interventional Neurology
Assistant Professor, Clinical Neurology and Neurosurgery
University of Miami and Jackson Memorial Hospital
Miami, FL

Osama O. Zaidat, MD, MS, FAAN, FAHA
St. Vincent Mercy Hospital
Toledo, OH

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2015

Ad.3 **Month and Year of most recent revision:** 08, 2015

Ad.4 **What is your frequency for review/update of this measure?** biannual

Ad.5 **When is the next scheduled review/update for this measure?** 02, 2016

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2866
Measure Title: CSTK-03: Severity Measurement Performed for Subarachnoid Hemorrhage (SAH) and Intracerebral Hemorrhage (ICH) Patients (Overall Rate)
Measure Steward: The Joint Commission
Brief Description of Measure: Proportion of SAH and ICH stroke patients age 18 years or older for whom a severity measurement (i.e., Hunt and Hess Scale for SAH patients or ICH Score for ICH patients) is performed prior to surgical intervention (e.g., clipping, coiling, or any surgical intervention) in patients undergoing surgical intervention and documented in the medical record; OR, documented within 6 hours of arrival at the hospital emergency department in patients who do not undergo surgical intervention.

This is the third measure in a set of measures developed for Joint Commission Comprehensive Stroke Certification. The other measures in the set include CSTK-01 National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients; CSTK-02 Modified Rankin Score (mRS) at 90 Days; CSTK-06 Nimodipine Treatment Administered. Although it is not required that these measures are reported in conjunction with each other, The Joint Commission develops measures in sets in order to provide as comprehensive a view of quality for a particular clinical topic as possible.
Developer Rationale: Subarachnoid hemorrhage (SAH) and intracerebral hemorrhage (ICH) are medical emergencies requiring rapid diagnosis and assessment. Early deterioration is common in the first few hours after onset, and associated with increased mortality rates of > 75% compared to 30-day mortality rates of 35%-52%. More than half of all deaths from these conditions occur within the first two days.

According to the American Heart Association/American Stroke Association, the severity of SAHs should be documented with the Hunt and Hess Scale (Connolly, 2012), and the severity of ICHs (Broderick, 2007) should be documented with ICH score to capture the clinical state of the patient. The severity of initial neurological injury should be determined and documented in the emergency department because it is a useful predictor of outcome and helpful in planning future care with family and physicians. For both severity methodologies, higher scores are associated with increased mortality.

Numerator Statement: CSTK-03 The number of SAH and ICH stroke patients for whom a severity measurement is performed prior to surgical intervention in patients undergoing surgical intervention and documented in the medical record; OR documented within 6 hours of arrival at the hospital emergency department in patients who do not undergo surgical intervention.

CSTK-03a The number of SAH stroke patients for whom a Hunt and Hess Scale is performed prior to surgical intervention in patients undergoing surgical intervention and documented in the medical record; OR documented within 6 hours of arrival at the hospital emergency department in patients who do not undergo surgical intervention.

CSTK-03b The number of ICH stroke patients for whom an ICH Score is performed prior to surgical intervention in patients undergoing surgical intervention and documented in the medical record; OR documented within 6 hours of arrival at the hospital emergency department in patients who do not undergo surgical intervention.
Denominator Statement: SAH and ICH stroke patients who arrive at this hospital emergency department (ED).

Denominator Exclusions: • Patients less than 18 years of age
• Patients who have a Length of Stay greater than 120 days
• Patients with Comfort Measures Only documented on the day of or day after hospital arrival
• Non-surgical patients discharged within 6 hours of arrival at this hospital
• Patients with admitting diagnosis of traumatic brain injury (TBI), unruptured arteriovenous malformation (AVM), and non-traumatic subdural hematoma (ICD-9-CM Other Diagnosis Codes as defined in Appendix A, Table 8.2f)

Measure Type: Process
Data Source: Electronic Clinical Data, Paper Medical Records
Level of Analysis: Facility, Population : National

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

Evidence Summary

As a process measure, there should be a determination if the measure focus is evidence-based, demonstrated via a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence that the measure focus leads to a desired health outcome. This new process measure assesses if a severity measurement (i.e., Hunt and Hess Scale for subarachnoid hemorrhage (SAH) patients and ICH Score for intracerebral hemorrhage (ICH) patients) was performed and documented prior to surgical intervention (e.g., clipping, coiling, or any surgical intervention), or performed within 6 hours of hospital arrival for SAH and ICH patients who do not undergo surgical intervention. The developer indicates the following:

- Performance of severity measurement for all SAH and ICH patients presenting to the hospital emergency department (ED) increases early detection and diagnosis of stroke and increases the identification of patients eligible for surgical intervention, i.e., clipping, coiling, and helps predict the prognosis for outcome.

The rationale for the measure is supported by three clinical guideline recommendations and include the following from Class I recommendations (Class I: Procedure/treatment SHOULD be performed/administered):

1. The initial clinical severity of a SAH should be determined rapidly by use of simple validated scales (e.g., Hunt and Hess, World Federation of Neurological Surgeons), because it is the most useful indicator of outcome after a SAH (*Level of Evidence: B*).
2. A baseline severity score should be performed as part of the initial evaluation of patients with ICH (*Level of Evidence: B*) (New recommendation).
3. ICH is a medical emergency, with frequent early, ongoing bleeding and progressive deterioration, severe clinical deficits, and subsequent high mortality and morbidity rates, and it should be promptly recognized and diagnosed (*Level of Evidence: A*).

Exception to evidence

N/A

Guidance from the Evidence Algorithm

Process measure/systematic review (Box 3) →QQC presented (Box 4) →Quantity: High; Quality: Moderate; Consistency: High →Box 5b Moderate (summarized as Moderate in the Evidence Form)

Questions for the Committee:

- *Does the Committee agree that the evidence provided (clinical guideline recommendations), systematic review of the evidence and other studies support the rationale for the measure and the relationship of the performance and documentation of severity assessment to patient outcomes?*
- *Is there evidence to support the 6 hour timeframe?*
  - *Is the evidence directly applicable to the process of care being measured?*
    -

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

Data on performance gap and opportunity improvement was provided from both the pilot test (66 sites and 2471 cases) and early 2015 data collected by the Joint Commission (varying numbers of sites and cases in 1st and 2nd quarter 2015). Assuming a target performance of 100%, the developer indicated there is possible gap of 80%. The pilot test hospital site scores ranged from 0 – 100%, with a mean of 19.63% on the overall measure score. Data from the first and second quarters of 2015 indicated higher mean performance for each component of this measure.

Q1: Overall national mean: .58; overall hospital mean: .59

Q2: Overall national mean: .62; overall hospital mean: .65

Disparities

As a new measure, detailed analysis and findings on disparities was not provided.

Questions for the Committee:

- *The mean performance in the field-test as compared to early 2015 data is significantly lower – does the distribution of performance scores between quartiles substantiate the need for a national performance measure?*
- *Is 100% the correct performance threshold?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **This is a 3-part process measure. The developers have provided evidence to support the measure in that the conditions (SAH and ICH) are significant medical emergencies with high morbidity and mortality, may be inappropriately diagnosed (and thus treated inappropriately), the often rapid changes in status and the fact that appropriate assessment of condition are associated with better outcomes support the measure. While as a process measure, documentation of patient status may or may not be associated with better outcomes; documentation may enhance the likelihood that status is indeed assessed.

**The measure focuses on time to performance of Hunt and Hess Scale in specific stroke populations. The scale provides a ranking of severity of the bleed with predictions as to outcomes. What I don't see is a documentation of relationship of performance of the assessment to subsequent treatment and outcome morbidities or mortality.

1b. Performance Gap

Comments: **Initial pilot data from 2012-3 indicated a potential 80% performance gap. Subsequent data from the 1st 2 quarters of

2015 after the Joint Commission implemented data collection for the 3 measure set showed much improved performance rates but I believe a significant performance gap of up to 40% persists.
**The data show a large gap between (80% - 40% is different studies) between performance of the assessment within the specified time and the desired 100%.

1c. High Priority (previously referred to as High Impact)

Comments: **The developers provide data to support their contention that the measures address a high priority aspect of healthcare including the significance of the conditions as a serious and prevalent health care issue and the economic and social burden of care and disability. The rationale for the proposed measures related to these facts is in the information the scales (Hunt & Hess for SAH & ICH Score for ICH) provide that inform prognosis and guide treatment.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): electronic clinical data, medical records

Specifications: Three component measure assessing if severity measurement was performed for an overall stroke population and then stratified for SAH and ICH patients prior to surgical intervention in patients undergoing surgical intervention and documented in the medical record or documented within 6 hours of arrival at hospital emergency department in patients who do not undergo surgical intervention. Data elements are clearly defined in the measure information form, and ICD-10 coding table/value set is provided. There is no risk adjustment or stratification based on population groups; however, the measure reports an overall rate and rates for SAH and ICH strokes to differentiate the type of severity assessment performed.

Questions for the Committee :

- *Are there any initial concerns about any of the data elements required to report this measure?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☒ **Data element** ☐ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing Reliability testing was performed at twelve participating pilot sites. Testing was conducted on a stratified random sample of records selected from each organization at the organization and measure category level. Hospitals were visited by teams of two Joint Commission staff during April, May, June, July, and August 2013. During these visits, Joint Commission staff re-abstracted records previously abstracted by hospital staff. A total of 281 records were re-abstracted. In cases of disagreement between the re-abstracted and original abstraction on a data element, reasons for disagreement were determined and adjudication was made as to whether original or re-abstraction findings were correct. Reliability was addressed by comparing the original abstracted and adjudicated re-

abstracted values, with the adjudicated value serving as the gold standard. The data analysis included both the percent agreement and the kappa statistic to adjust for chance agreement for categorical data elements.

Results of reliability testing Percent agreement/match of data elements between the abstractors ranged from a low of 71.5% for discharge time and a high of 99.3% for confirmation of Emergency department patient. Kappas were calculated on three data elements: Initial Hunt and Hess performed (K=0.91), Initial ICH Score performed (K=0.86) and Emergency Department patient (K=0.96). These scores suggest almost perfect agreement based on the Landis and Koch classification.

Guidance from the Reliability Algorithm Precise specifications (Box 1) →empirical testing with statistical tests as specified (Box 2) → reliability testing with patient-level data elements (Box 8) →assessed all critical data elements for numerator and one of the exclusions, with the remaining denominator data elements obtained via administrative data→ high/moderate certainty the data elements are reliable (Box 10a) → Moderate

Questions for the Committee:

- *Does the lower percent agreement (71.5%) as compared to the other data elements cause an concern for reliability?*
- *The measure was tested with ICD-9 data to identify stroke patients, does the Committee anticipate reliability to be impacted with the switch to ICD-10?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☐ **Yes** ☒ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING
Validity testing level ☐ **Measure score** ☒ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity only**
- ☒ **Empirical validity testing of the measure score**

Validity testing method: Face Validity:

- Face validity of data elements was reported but did not address face validity of the measure score as a representation of quality, as required for the criterion.
- Data element validity was assessed for accuracy and clarity by hospitals, but the data element validity criterion requires comparison to a gold standard.

Validity testing results:

CSTK03 and CSTK06 are measures for hemorrhagic stroke. The measure rate among these measures are expected to be correlated. The Pearson correlation coefficient is calculated.

Pearson Correlation Coefficient interpretation: the range is from -1 to 1. When the correlation coefficient is close to +1 or -1, it means that there is strong correlation; p value is utilized to determine if the corrlation coefficient is significant or not. If it is less than 0.05, then the conclusion usually is significant.

The table below demonstrates that CSTK03 rate was negtively and slightly correlated to rest of these measures. Because the data is not normally distributed, the corrlation coefficient on correlation may not be the best measure of association. Also the large proportion of zero measure rates for some of the measures, while demonstrating considereable room for improvement, would tend to skew the correlations.

Pearson Correlation Coefficients, N = 66 Prob > |r| under H0: Rho=0

	CSTK03	**CSTK06**
Correlation Coefficient	1.00000	-0.02229
P value		0.8590

The average rating for measure CSTK-03 numerator and denominator statements, including the clarity of numerator and denominator inclusions and exclusions was 3.74. Although this rating is slightly lower than those for other measures in the CSTK set, it can still be used to distinguish good and poor quality.

The developers indicate, the rating for data availability and accessibility reflects that the ICH Score was not always documented in the medical record; however, documentation of the Hunt and Hess Scale was consistent and not an issue. Additionally, ICD-9-CM Other Procedure Times were often missing, although the ICD-9-CM Principal Procedure Time was consistently documented. We expect that data availability will improve with use over time.

1Q and 2Q 2015 Data:
Overall the positive inter-correlations indicates convergent validity of all the measures.
They are positively correlated with other evidence-based processes of care.

Questions for the Committee:

- *Do you agree the test sample is adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made? Does the lower rating on data availability and accessibility for the ICH score cause concern for the overall metric?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*
- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

There were 3830 admissions included in the initial cohort and diagnosed hemorrhagic stroke. From among the 3830 admissions in 66 hospitals, the descriptive statistics are given below.
Exclusion: Non-surgical patients discharged within 6 hours of arrival at this hospital
Overall Occurrence n = 9
Overall Occurrence Percentage: 0.23%

Exclusion: Patients with admitting diagnosis of traumatic brain injury (TBI), unruptured arteriovenous malformation (AVM), and non-traumatic subdural hematoma
Overall Occurrence n = 15
Overall Occurrence Percentage: 0.39%

1Q and 2Q 2015 Data:

There were 65,389 admissions selected from the initial cohort. From among the 65,389 admissions in 42 hospitals, the descriptive statistics are given below.
Applied To Measure CSTK-01 CSTK-03 CSTK-06
Exclusion: Comfort Measures - 1 Day 0 or 1
Overall Occurrence n = 1,300
Overall Occurrence Percentage: 2%
Minimum: 0.47%
Median: 3%
Maximum: 6%

Applied To Measure CSTK-3
Exclusion: Non-surgical patients discharged within 6 hours of arrival at this hospital
Overall Occurrence n = 20,266
Overall Occurrence Percentage: 62%
Minimum: 47%
Median: 54%
Maximum: 17.39%
NMISS= 20,232 missing observation

Applied To Measure CSTK-3
Exclusion: Patients with admitting diagnosis of traumatic brain injury (TBI), unruptured arteriovenous malformation (AVM), and non-traumatic subdural hematoma (ICD-9-CM Other Diagnosis Codes as defined in Appendix A, Table 8.2f)
None

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

o

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

1Q and 2Q 2015 Data:
Overall descriptive statistics for CSTK selected measures: N = 42 certified comprehensive stroke hospitals: n = 65,389
Results for CSTK-03 were:

CSTK-03
Median: 61%
Percentile 10%: 33%
Percentile 25%: 48%
Percentile 75%: 79%
Percentile 90%: 89%

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

Guidance from algorithm: Specifications consistent with evidence (Box 1) → potential threats to validity mostly assessed (Box2) →empirical validity testing (Box 3) → measure score testing (Box 6) → sound conceptualization and hypotheses (Box 7) →high confidence that score are a valid indicator of quality(Box 8a) → High

Preliminary rating for validity: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **The measure specifications appear to be consistent with the evidence presented by the developers.

**This appears to me to be a significant question - as alluded to in 1a. It is unclear what the relationship of passing vs. not passing this test is to changes in morbidity, mortality or increased length of survival, or time interval to a subsequent event.

**Although there are many data elements to capture, all are well defined and seem feasible.

2a2. Reliability Testing

Comments: **Face valid was assesses with surveys and focus groups. Convergent validity was established with patient data from the Joint Commission warehouse.

Correlations btw the the process measures were established...although I'm not entirely clear why?

**Again, numbers are sufficient; however, relationship of the data to outcomes that would matter to patients and families is not presented.

*Reliability testing was done at 12 pilot sites with a stratified sample of records (n=281). Testing was done by comparing raters assessment with an adjudicated score...this seems a little weird to me, but I'm not sure how else this could have been done. The % agreement was 71-99%.

2b2. Validity Testing

Comments: **I don't believe so. There is some discussion about ICH scores but the state that they "expect that data availability will improve with use over time". Files with missing data would not be used to process the measure.

**Exclusions seem appropriate.

**Face valid was assessed with surveys and focus groups. Convergent validity was established with patient data from The Joint Commission warehouse. Correlations btw the the process measures were established…although I'm not entirely clear why

2b2. Validity Testing

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **Reliability testing was done at 12 pilot sites with a stratified sample of records (n=281). Testing was done by comparing raters assessment with an adjudicated score...this seems a little weird to me, but I'm not sure how else this could have been done. The % agreement was 71-99%.

**Data seem sufficient for a reliability assessment.

**I don't believe so. There is some discussion about ICH scores but they state that they "expect that data availability will improve with use over time". Files with missing data would not be used to process the measure

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

This measure is one in a set of measures implemented with discharges effective January 1, 2015 for Joint Commission Comprehensive Stroke Certification. There are no fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain. The measure is collected either through electronic medical record or paper medical record abstraction. Not all hospitals have fully implemented EMRs and thus there may be additional burden for those hospitals. Based on the testing results, there were some data elements with lower completeness ratings that should be followed for improvement in documentation to be sure missing data do not introduce bias.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **This measure is part of a measures set implemented with discharges effective January 1, 2015 for Joint Commission Comprehensive Stroke Certification. There are no fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain. The measure is collected either through electronic medical record or paper medical record abstraction. Not all hospitals have fully implemented EMRs and thus there may be additional burden for those hospitals. Based on the testing results, there were some data elements with lower completeness ratings that should be followed for improvement in documentation to be sure missing data do not introduce bias.

**The value of this measure (to me) is the relationship of the cost of training, administration and data collection to the improved outcomes (decreased morbidity, mortality, recovery period, life extension, time to subsequent event, etc.). While collection of these data is relatively simple and straightforward (if the data are entered) the value of the data relative to the cost of collection, etc. is unclear.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure [from OPUS]

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details N/A – will be used in Joint

Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint

Commission
• Purpose: A certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

Improvement results
The initial performance measure gap for this measure was huge at ~80%. Although considerable improvement has been noted since the pilot test of the measure based on the first two quarters of Joint Commission ORYX performance measurement data, the performance gap remains significant, especially for ICH patients. For 2Q2015 (N=51 hospitals), a wide range of performance was noted with a 10% gap at the 90th percentile and 70% gap at the 10th percentile.

Unexpected findings (positive or negative) during implementation N/A new measure

Potential harms None identified

Feedback :
This measure was collected during a 6 month pilot process for Joint Commission Comprehensive Stroke Certification. It was learned via discussions and written pilot evaluation materials that there were some issues related to data abstraction for the following data elements:
• Initial Hunt and Hess Scale Performed
o Clarification was requested regarding the use of abbreviations and alternative terms for the Hunt and Hess Scale.
• Initial ICH Score Performed
o Clarification was requested regarding the components of the ICH score and how to calculate the ICH score.

***Questions for the Committee**:*
- o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 4: Usability and Use

4a. Accountability and Transparency
4b. Improvement
4c. Unintended Consequences
Comments: **The measures are currently being used in the Disease-Specific Care Certification for Comprehensive Stroke Centers. Performance data shows improvement in the first 2 quarters of 2015 over earlier data, but I believe there is still a need for improvement. The developers cite some issues with the Comprehensive Stroke Certification data collection (abbreviations, alternative terms, calculation of the ICH score) which merit discussion and an action plan. Addressing these issues may improve overall performance data.
**Assuming there are data that establishes a clear improvement in outcome from timely use of this measure, the usability is clear. This measure does appear to show advantages over other measures and does appear harmonized with the other Joint Commission related measures.

Criterion 5: Related and Competing Measures

Related or competing measures
Related measures:
- CSTK-01 National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients;
- CSTK-02: Modified Rankin Score (mRS) at 90 Days
- CSTK-06: Nimodipine Treatment Administered

Harmonization

- All related measures are from the same developer and are harmonized**.**

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Title: CSTK-03: Severity Measurement Performed for SAH and ICH Patients (Overall Rate)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure title

Date of Submission: 1/15/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

Subcriterion 1a. Evidence to Support the Measure Focus

The measure focus is a health outcome or is evidence-based, demonstrated as follows:

- Health outcome:[3] a rationale supports the relationship of the health outcome to processes or structures of care.
- Intermediate clinical outcome, Process,[4] or Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence[5] that the measure focus leads to a desired health outcome.
- Patient experience with care: evidence that the measured aspects of care are those valued by patients and for which the patient is the best and/or only source of information OR that patient experience with care is correlated with desired outcomes.
- Efficiency:[6] evidence for the quality component as noted above.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement.

5. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
6. Measures of efficiency combine the concepts of resource use and quality (NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of:

Outcome

☐ Health outcome: Click here to name the health outcome

Health outcome includes patient-reported outcomes (PRO, i.e., HRQoL/functional status, symptom/burden, experience with care, health-related behaviors)

☐ Intermediate clinical outcome: Click here to name the intermediate outcome

☐ Process: Severity measurement (i.e., Hunt and Hess Scale for subarachnoid hemorrhage (SAH) patients and ICH Score for intracerebral hemorrhage (ICH) patients) performed and documented prior to surgical intervention (e.g., clipping, coiling, or any surgical intervention), or performed within 6 hours of hospital arrival for SAH and ICH patients who do not undergo surgical intervention.

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME PERFORMANCE MEASURE *If not a health outcome, skip to 1a.3*

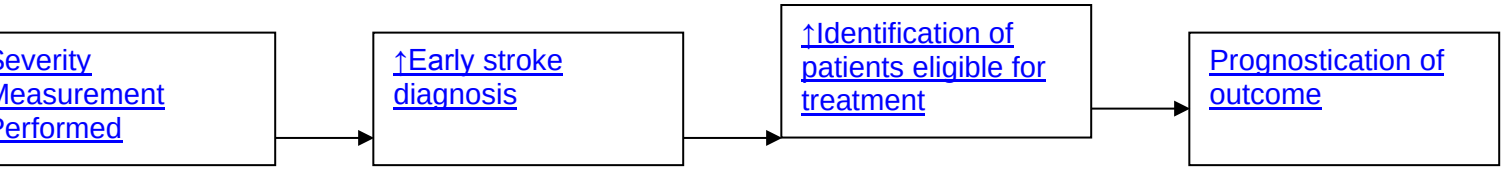
1a.2. Briefly state or diagram the linkage between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

Note: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the linkages between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.



Performance of severity measurement for all SAH and ICH patients presenting to the hospital emergency department (ED) increases early detection and diagnosis of stroke and increases the identification of patients eligible for surgical intervention, i.e., clipping, coiling, and helps predict the prognosis for outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☐ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

Connolly ES Jr, Rabinstein AA, Carhuapoma JR, Derdeyn CP, Dion J, Higashida RT, Hoh BL, Kirkness CJ, Naidech AM, Ogilvy CS, Patel AB, Thompson G, Vespa P, on belhalf of the American Heart Association Stroke Council, Council on Cardiovascular Radiology and Intervention, Council on Cardiovascular Nursing, Council on Cardiovascular Surgery and Anesthesia, Council on Clinical Cardiology. Guidelines for the management of aneurysmal subarachnoid hemorrhage: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke.* 2012;43:1-27.

URL:

http://stroke.ahajournals.org/content/early/2012/05/03/STR.0b013e3182587839.full.pdf

Broderick J, Connolly S, Feldmann E, Hanley D, Kase C, Krieger D, Mayberg M, Morgenstern L, Ogilvy CS, Vespa P, Zuccarello M. Guidelines for the Management of Spontaneous Intracerebral Hemorrhage: A Guideline from the American Heart Association/American Stroke Association. *Stroke.* 2007;38:2001-2023.

URL:

http://stroke.ahajournals.org/content/38/6/2001.full.pdf

Hemphill CJ III, Greenberg SM, Anderson CS, Becker K, Bendok CR, Cushman M, Fung GL, Goldstein JN, Macdonald RL, Mitchell PH, Scott PA, Selim MH, Woo D. Guidelines for the Management of Spontaneous Intracerebral Hemorrhage: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association. *Stroke.* 2015;46:1-29.

https://stroke.ahajournals.org/content/early/2015/05/28/STR.0000000000000069.full.pdf+html

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

Connolly, 2012. Natural History and Outcome of aSAH (Page 8)

Class I

4. The initial clinical severity of aSAH should be determined rapidly by use of simple validated scales (e.g., Hunt and Hess, World Federation of Neurological Surgeons), because it is the most useful indicator of outcome after aSAH (*Level of Evidence: B*).

Hemphill, 2015. Emergency Diagnosis and Assessment (Page 4)

Class I

5. A baseline severity score should be performed as part of the initial evaluation of patients with ICH (*Level of Evidence: B*) (New recommendation).

Broderick, 2007. Emergency Diagnosis and Assessment of ICH and Its Causes (Page 2004)

Class I

1. ICH is a medical emergency, with frequent early, ongoing bleeding and progressive deterioration, severe clinical deficits, and subsequent high mortality and morbidity rates, and it should be promptly recognized and diagnosed (*Level of Evidence: A*).

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

Class I: Procedure/treatment SHOULD be performed/administered.

Level A: Data derived from multiple randomized clinical trials or meta-analyses.

Level of Evidence B: Data derived from a single randomized trial or nonrandomized studies.

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

ACCF/AHA Classification of Recommendation and Level of Evidence

Classification Types

Class II a: It is reasonable to perform procedure/administer treatment.

Class II b: Procedure/Treatment may be considered.

Class III: Procedure/Treatment not helpful/no proven benefit/may be harmful.

Level of Evidence

Level C: Only consensus opinion of experts, case studies, or standard of care.

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

Same as 1a.4.1

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

Rosen DS, Macdonald RL. Subarachnoid Hemorrhage Grading Scales: A systematic Review. *Neurocritical Care.* 2005;2:110-118.

URL:

http://www.cmp-manual.cz/skaly/subarachnoid_hemorrhage_grading_scales.pdf

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Same as 1a.6.1

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

The information in the following questions in this section is based upon the systematic review conducted by Rosen and Macdonald in 2005.

The review included articles pertaining to cerebral aneurysm if the purpose of the article was to use or create a grading scale on which to predict outcome after SAH. Articles were excluded if they examined a single prognostic factor and related this factor to outcome or if factors were used to predict some other complication of SAH, such as vasospasm, rather than outcome. The following four grading systems were compared: Hunt and Hess Scale, Fisher Scale, Glasgow Coma Scale (GCS), and World Federation of Neurological Surgeons Scale (WFNS).

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

An overall grade for the quality of quoted evidence was not assigned for either review.

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

An overall grade for the quality of quoted evidence was not assigned for either review.

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

1966-March 2004

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

The quantity of evidence was not specified in the review.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Overall quality of evidence appears to be moderate. The literature addressing SAH grading scales contains some deficiencies. Most of the reported grading systems are based on expert opinions of the authors and are applied to a relatively small set of patients, usually selected from a single institution. No prospective, controlled, comparison studies have been published.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

The *K* values for the components of the Hunt and Hess Scale range between 0.25 for headache and 0.52 for level of consciousness. The overall *K* for the Hunt and Hess Scale is 0.42, which is significantly greater than that expected by chance.

Comparison of the Hunt and Hess Scale, GCS, and WFNS Scale in a series of 185 patients with SAH showed that the Hunt and Hess Scale has the strongest predictive power for outcome at 6 months, as assessed by the Glasgow Outcome Scale (GOS). Scores on the day of surgery were also found to be of more prognostic value than values observed immediately after hospitalization.

The most important advantages of the Hunt and Hess Scale are that is widely known in the neuroscientific community and that it is well-entrenched in the literature on SAH. A review of 184 articles regarding SAH published in nine neurological journals between 1985 and 1992 showed that 71% of authors used the Hunt and Hess and Kosnik Scales, 19% used the WFNS Scale or GCS, and

10% used other scales. Another advantage is its ease of administration because multiple steps are not required to derive a comprehensive grade.

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

The evidence indicates that the Hunt and Hess Scale is a strong predictor of outcome after SAH. Variable interpretation is an inherent harm. Many of the terms used to define the grades of the Hunt and Hess scale (drowsy, stupor, deep coma) are somewhat vague. Ambiguity between grades may reduce inter-rater reliability of the scale; however, the benefits of the Hunt and Hess Scale's predictive power and acceptance outweigh the harm.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

Not applicable

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

A literature search of Medline, PubMed, CINAHL, Cochrane Database of Systematic Reviews, The National Guideline Clearinghouse, The National Quality Measure Clearinghouse, and other sources was conducted in February 2012; from this search eight references were cited for use as evidence for the CSTK-03 Pilot measure. These references included evidence-based guidelines, randomized control trials, cohort studies, and other well-referenced publications. In June 2014, an additional literature search via PubMed, Cochrane Database of Systematic Reviews, Medline, Google, EBSCO, and Research Gate was conducted and yielded seven additional references for ICH Score, including the cohort studies summarized below.

1a.8.2. Provide the citation and summary for each piece of evidence.

Clarke JL, Johnston SC, Farrant M, Bernstein R, Tong D, Hemphill JC III. External validation of the ICH score. *Neurocritical Care.* 2004;1:53-60.

Summary:

The purpose of this study was to determine whether the ICH Score accurately risk-stratifies patients with acute intracerebral hemorrhage (ICH). An independent cohort of 175 patients presenting with acute ICH to Stanford Medical Center and Santa Clara Valley Medical Center from 1998 to 2000 and assigned a ICD-9-CM Principal Diagnosis Code of 431 Intracerebral Hemorrhage at discharge were reviewed. Outcome was assessed as mortality at 30 days after ICH. Overall 30-day mortality was 40% (n=70). ICH Scores ranged from 0 to 5, and each increase in the ICH score was associated with an increase in 30 day mortality ($p < 0.01$ for trend). The study concluded that the ICH score accurately stratifies outcome in an external patient cohort and is a validated clinical grading scale that can be easily and rapidly applied at the time of ICH presentation.

Hemphill JC III, Farrant M, Neill TA. Prospective valiation of the ICH Score for 12-month functional outcome. *Neurology.* 2009;73:1088-1094.

Summary:

The purpose of this prospective longitudinal observational study was to determine whether the ICH Score reliably stratifies patients with acute ICH with regard to 12-month functional outcome as assessed by the modified Rankin Scale (mRS). A total of 243 patients with acute nontraumatic ICH who presented to the emergency department (ED) of San Francisco General Hospital (SFGH) or UCSF Medical Center from June 1, 2001 through May 31, 2004 were followed for up to 12 months. Clinical status using the the mRS was assessed at discharge, 30 days, 3, 6, and 12 months. The study demonstrated that the ICH Score is a valid clinical grading scale for stratifying the likelihood of favorable functional outcome through the first year after acute ICH, out to 12 months. Importantly, in this study the ICH Score was valid and significant regardless of the specific outcome cutpoint chosen to define favorable functional outcome ($p < 0.05$, test for trend).

Bulic S, Sanossian N, Liebeskind D, Villablanca P, Hamilton S, Conwit R, Savers J. Validation of the ICH score in Hyperacute Intracerebral Hemorrhage. *Neurology.* 2014;82(10):S25.004.

Summary:

The NIH Field Administration of Stroke Therapy magnesium (FAST-MAG) data based was used to identify 364 consecutive patients with spontaneous ICH on initial imaging. Cases were reviewed to validate the ICH Score in a modern cohort of ICH patients presenting to the emergency department within 2 hours from symptom onset. The study concluded that in the hyperacute setting, all patients with ICH Score $\geq$ 4 had poor outcome, defined as modified Rankin Score (mRS) 4-6 and death as mRS 6 at 3 months. ICH Score predicted mortality and poor outcome ($p<0.0001$).

Wang W, Lu J, Wang C, Wang Y, Li H, Zhao X on belhalf of the investigators for the China National Stroke Registry (CNSR) Investigators. Prognostic value of the ICH score and ICH-GS score I

Chinese intracerebral hemorrhage patients: analysis from the China National Stroke Registry (CNSR). *Plos ONE.* 2013;8(10):e77421.

Summary:

In this multicenter, prospective cohort study of 3,255 ICH patients, both the ICH Score and ICH-GS Score were found to be effective for predicting short-term and long-term favorable functional outcome after ICH. ICH accounts for 55% of all strokes in China, a much higher percentage than the 10%-15% noted in Western populations. This study assessed functional outcome at discharge, 3, 6, and 12 months after discharge using the ICH Score and ICH-GS Score. The values of the Area Under Curve (AUC) for the ICH score were 0.72, 0.76, 0.76, and 0.75 respectively. The ICH-GS score was a better predictor at 6 and 12 months (i.e., 0.71, 0.77, 0.78, 0.78).

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
CSTK-03_Measure__Evidence_6.5.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Subarachnoid hemorrhage (SAH) and intracerebral hemorrhage (ICH) are medical emergencies requiring rapid diagnosis and assessment. Early deterioration is common in the first few hours after onset, and associated with increased mortality rates of > 75% compared to 30-day mortality rates of 35%-52%. More than half of all deaths from these conditions occur within the first two days.

According to the American Heart Association/American Stroke Association, the severity of SAHs should be documented with the Hunt and Hess Scale (Connolly, 2012), and the severity of ICHs (Broderick, 2007) should be documented with ICH score to capture the clinical state of the patient. The severity of initial neurological injury should be determined and documented in the emergency department because it is a useful predictor of outcome and helpful in planning future care with family and physicians. For both severity methodologies, higher scores are associated with increased mortality.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.** *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Pilot Test Findings:
During the six-month pilot test (October 1, 2012-March 31, 2013), sixty-six sites submitted data for 10,218 completed patient records. For this measure, 2471 cases were assigned an ICD-9-CM Principal Diagnosis Code of 430 Subarachnoid Hemorrhage or 431 Intracerebral Hemorrhage at discharge and captured in the denominator population, and 485 of these cases met the numerator requirements. The performance rates varied widely across sites for this measure with results ranging from a low of 0% to a high of 100%. The average rate for all sites collecting data for this measure was 19.63%, indicating a potential performance gap of approximately 80% if the optimal rate is 100%.

In January, 2015, The Joint Commission implemented data collection for the comprehensive stroke (CSTK) measure set to meet performance measurement requirements for its Comprehensive Stroke Certification Program. Below is the specified level of analysis for CSTK-03 Severity Measurement for SAH and ICH Patients for the two quarters of data received to date for this measure.

1Q 2015 (0verall Rate): 1786 denominator cases; 1042 numerator cases; 39 hospitals; 0.58343 national aggregate rate; 0.58578 mean of hospital rates; 0.2295 standard deviation; 0.86057 90th percentile rate; 0.78049 75th percentile rate/upper quartile; 0.58333 50th percentile rate/median rate; 0.42857 25th percentile rate/lower quartile; and, 0.23684 10th percentile rate.

1Q 2015 (SAH): 578 denominator cases; 362 numerator cases; 39 hospitals; 0.6263 national aggregate rate; 0.63778 mean of hospital rates; 0.25176 standard deviation; 0.95455 90th percentile rate; 0.8 75th percentile rate/upper quartile; 0.66667 50th percentile rate/median rate; 0.5 25th percentile rate/lower quartile; and, 0.22222 10th percentile rate.

1Q 2015 (ICH): 1208 denominator cases; 680 numerator cases; 39 hospitals; 0.56291 national aggregate rate; 0.55635 mean of hospital rates; 0.2585 standard deviation; 0.88889 90th percentile rate; 0.76667 75th percentile rate/upper quartile; 0.55 50th percentile rate/median rate; 0.42424 25th percentile rate/lower quartile; and, 0.2 10th percentile rate.

2Q 2015 (Overall Rate): 2310 denominator cases; 1443 numerator cases; 51 hospitals; 0.62468 national aggregate rate; 0.65156 mean of hospital rates; 0.23596 standard deviation; 0.90323 90th percentile rate; 0.82432 75th percentile rate/upper quartile; 0.73913 50th percentile rate/median rate; 0.4901 25th percentile rate/lower quartile; and, 0.2931 10th percentile rate.

2Q 2015 (SAH): 854 denominator cases; 572 numerator cases; 51 hospitals; 0.66979 national aggregate rate; 0.67815 mean of hospital rates; 0.25727 standard deviation; 1.0 90th percentile rate; 0.88 75th percentile rate/upper quartile; 0.72727 50th percentile rate/median rate; 0.52174 25th percentile rate/lower quartile; and, 0.33333 10th percentile rate.

2Q 2015 (ICH): 1456 denominator cases; 871 numerator cases; 51 hospitals; 0.59821 national aggregate rate; 0.62517 mean of hospital rates; 0.26669 standard deviation; 0.875 90th percentile rate; 0.83333 75th percentile rate/upper quartile; 0.725 50th percentile rate/median rate; 0.42857 25th percentile rate/lower quartile; and, 0.25 10th percentile rate.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

Citation:
Brain Aneurysm Foundation. Understanding: Brain Aneurysm Statistics and Facts.
http:// www.bafound.org/Statistics_and_Facts. Published 2014. Accessed June 5, 2014.

Summary:
Accurate early diagnosis is critical, as the initial hemorrhage may be fatal, may result in devastating neurologic outcomes, or may produce minor symptoms. Despite widespread neuroimaging availability, misdiagnosis or delays in diagnosis occurs in up to 25% of patients with subarachnoid hemorrhage (SAH) when initially presenting for medical treatment. Failure to do a scan results in 73% of these misdiagnoses. This makes SAH a low-frequency, high-risk disease.

Citation:
Mayer PL, Awad IA, Todor R, et. al. Misdiagnosis of symptomatic cerebral aneurysm: prevalence and correlation with outcome at four institutions. Stroke. 1996;27:1558-1563.

Summary:
Data analyzed: 1990's; 4 Connecticut, U.S. neurosurgical units; 217 SAH patients. Fifty-four (25%) of patients with subarachnoid hemorrhage initially received an incorrect diagnosis; most of them were in good clinical condition at presentation. The condition of the 54 patients with incorrect diagnosis worsened, usually as a result of recurrent bleeding, before definitive treatment was begun. Of the 163 patients given a correct diagnosis, the condition of only 2.5% worsened.

Citation:
Kowalski RG, Claassen J, Kreiter KT, Bates JE, Ostapkovich ND, Connolly ES, Mayer SA. Initial misdiagnosis and outcome after subarachnoid hemorrhage. JAMA. 2004 Feb 18;291(7): 866-869.

Summary:
Data analyzed: August 1996 – August 2001; U.S. tertiary care urban hospital; 482 SAH patients. Misdiagnosis of SAH occurred in 12% of patients and was associated with a smaller hemorrhage and normal mental status. Among individuals who initially present in good condition, misdiagnosis is associated with increased mortality and morbidity.

Citation:
Vermeulen MJ, Schull MJ. Missed diagnosis of subarachnoid hemorrhage in the emergency department. Stroke. 2007;38:1216-1221.

Summary:
Data analyzed: April 1, 2002 – March 31, 2005; 176 Canadian hospitals with emergency departments (EDs); 1603 patients hospitalized with a diagnosis of nontraumatic SAH. Among all nontraumatic SAH patients admitted to an Ontario hospitals, 5.4% had been misdiagnosed on a prior visit to an ED. About 1 in 20 persons with SAH are missed on their first presentation to an ED, and the risk is greater in patients with low acuity presentations.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

This is the initial submission of this measure.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.

Brain Aneurysm Foundation. Understanding: Brain Aneurysm Statistics and Facts. http:// www.bafound.org/Statistics_and_Facts. Published 2014. Accessed June 5, 2014.

Ruptured brain aneurysms account for 3-5% of all new strokes. Women compared to men suffer from brain aneurysms at a ratio of 3:2.

_____

Morgenstern LB, Smith MA, Lisabeth LD, Risser JM, Uchino K, Garcia N, Longwell PJ, McFarling DA, Akuwumi O, Al-Wabil A, Al-Senani F, Brown DL, Moyé LA. Excess Stroke in Mexican Americans compared with non-Hispanic Whites: the Brain Attack Surveillance in Corpus Christi Project. Am J Epidemiol. 2004 Aug 15;160(4):376-383.

The project studied 2,350 cerebrovascular events occurring from January 2000 to December 2002 in Nueces County, Texas. Of the completed strokes, 53% were in Mexican Americans. The crude cumulative incidence was 168/10,000 in Mexican Americans and 136/10,000 in non-Hispanic Whites. Mexican Americans experienced a higher incidence of intracerebral hemorrhage and subarachnoid hemorrhage than non-Hispanic whites, adjusted for age. Intracerebral hemorrhage was more common in Mexican Americans (age-adjusted risk ratio = 1.63, 95% confidence interval: 1.24, 2.16). The subarachnoid risk age-adjusted risk ratio was 1.57 (95% confidence interval: 0.86, 2.89).

_____

Schievink WI, Riedinger M, Jhutty TK, Simon P. Racial disparities in subarachnoid hemorrhage mortality: Los Angeles County, California, 1985-1998. Neuroepidemiology. 2004 Nov-Dec;23(6):299-305.

The number of SAH deaths was 2,897. The age-adjusted SAH mortality rate was 1.9 in whites, 2.7 in Hispanics, 3.0 in Asians and 3.7 in blacks. In those younger than 70 years of age, the SAH mortality rate among blacks was 2.2 times that of whites and 1.8 times that of Hispanics and Asians. The SAH mortality rate declined after age 70 in blacks. The SAH mortality rate was higher in women than in men in all races and it was highest in elderly Asian women (23.5 per 100,000). An inverse relationship was observed between income and SAH mortality rates in all racial groups except whites.

_____

Jaja BNR, Saposnik G, Nisenbaum R, Lo BWY, Schweizer TA, Thorpe KE, Macdonald RL. Racial/ethnic differences in outcomes following subarachnoid hemorrhage. 2013 Sep 10; DOI: 10.3171/2013.7.JNS13544.

Using 2005-2010 data from the Nationwide Inpatient Sample, Jaja and colleagues conducted a cross-sectional study of hospital discharges for patients whose principal diagnosis was SAH unrelated to trauma (n=31,631), In this study, inpatient mortality was the primary outcome and discharge to institutional care was the secondary outcome. The researchers found a crude inpatient mortality rate of 22% and a 42% rate of hospital discharge to institutional care. Multivariable analyses identified race/ethnicity as a significant predictor of both inpatient mortality (p=0.003) and discharge to institutional care (p < 0001). Hispanic, black, Native American, and Asian/Pacific Islander patients were compared to whites. The study found that Hispanic patients were the least likely to have a poor outcome, and Asian/Pacific Islander patients were most likely to have a poor outcome.

Stansbury JP, Jia H, Williams LS, Vogel WB, Duncan PW. Ethnic disparities in stroke: epidemiology, acute care, and postacute outcomes. Stroke. 2005;36:374-386.

In Stansbury's, et. al., analysis of current literature, it was noted that the mortality rates with ICH for blacks and Asian/Pacific Islanders were 1.7- and 1.5-times those of whites in 1995 to 1998 data, and the rates from SAH were higher for all ethnic minorities. Black stroke mortality was the highest across all stroke types in all but the elderliest age groups.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a

substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
A leading cause of morbidity/mortality, High resource use, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Stroke ranks as the number four cause of death in the United States, following diseases of the heart, cancer, and chronic lung-related diseases. Each year, ~ 795,000 people experience a new or recurrent stroke. Approximately 610,000 of these are first attacks, and 185,000 are recurrent strokes. These numbers equate to one stroke victim every 40 seconds on average. Intracerebral hemorrhage (ICH) is the second most common type of stroke, accounting for 10% of all strokes. While subarachnoid hemorrhage (SAH) represents only 3% of all strokes, it is the most deadly type of stroke resulting in more than a 50% mortality rate. Of SAH survivors, approximately half will suffer a permanent disability.

The mean lifetime cost of ischemic stroke, including inpatient care, rehabilitation, and follow-up as necessary for residual deficits are estimated at $140,048 per person. Death within 7 days, SAH, and stroke while hospitalized for another condition are associated with higher costs.

A large amount of work has been devoted to the development of grading scales to aid the assessment of stroke. The Hunt and Hess Scale and ICH score are two such scales. Both scales measure the severity of initial neurological injury specific for the type of stroke, (i.e. Hunt and Hess for SAH; ICH Score for ICH). These scales provide prognostic information regarding outcome, guide treatment decisions, and can be used to standardize patient assessment across medical centers.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**
Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, Huffman MD, Judd SE, Kissela BM, Lackland DT, Lichtman JH, Lisabeth LD, Simin L, Mackey RH, matchar DB, McGuire DK, Mohler ER III, Moy CS, Muntner P, Mussolino ME, Nasir K, Neumar RW, Graham N, Palaniappan L, Pandey DK, Reeves MJ, Rodriguez CJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Willey JZ, Woo D, Yeh RW, Turner JZ. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation. 2015;131:e-151-e154.

Rosen DS, Macdonald RL. Subarachnoid hemorrhage grading scales: a systematic review. Neurocritical Care. 2005;2:110-118.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*
Not applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Functional Status, Health and Functional Status : Functional Status

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to*

general information.)
http://www.jointcommission.org/assets/1/6/CSTKManual2015August.pdf

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Copy_of_AppendixACSTKTables_ICD10codes.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
Not Applicable

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
CSTK-03 The number of SAH and ICH stroke patients for whom a severity measurement is performed prior to surgical intervention in patients undergoing surgical intervention and documented in the medical record; OR documented within 6 hours of arrival at the hospital emergency department in patients who do not undergo surgical intervention.

CSTK-03a The number of SAH stroke patients for whom a Hunt and Hess Scale is performed prior to surgical intervention in patients undergoing surgical intervention and documented in the medical record; OR documented within 6 hours of arrival at the hospital emergency department in patients who do not undergo surgical intervention.

CSTK-03b The number of ICH stroke patients for whom an ICH Score is performed prior to surgical intervention in patients undergoing surgical intervention and documented in the medical record; OR documented within 6 hours of arrival at the hospital emergency department in patients who do not undergo surgical intervention.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Episode of care

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Twelve data elements are used to calculate the numerator. Data elements and definitions:
• Arrival Date - The earliest documented month, day, and year, the patient arrived at the hospital.

• Arrival Time - The earliest documented time (military time) the patient arrived at the hospital.

• ICD-10-PCS Other Procedure Code Date – The month, date, and year when the other procedure(s) was performed.

• ICD-10-PCS Other Procedure Code Time - The time (military time) when the other procedure(s) was performed.

• ICD-10-PCS Principal Procedure Code Date – The month, date, and year when the principal procedure was performed.

• ICD-10-PCS Principal Procedure Code Time - The time (military time) when the principal procedure was performed.

• Initial Hunt and Hess Scale Date - The month, date, and year the Hunt and Hess Scale was first performed at the hospital.

• Initial Hunt and Hess Scale Performed - Documentation of the first Hunt and Hess Scale that was done at this hospital. The Hunt

and Hess Scale is a grading system used to classify the severity of a subarachnoid hemorrhage based on the patient's clinical condition. The scale ranges from a score of 1 to 5. It is used as a predictor of prognosis and outcome with a higher grade correlating to a lower survival rate. Allowable Values: Yes or No/UTD.

• Initial Hunt and Hess Scale Time - The time (military time) for which the Hunt and Hess Scale was first documented at the hospital.

• Initial ICH Score Date - The month, date, and year the ICH Score was first performed at the hospital.

• Initial ICH Score Performed - Documentation of the first ICH Score that was done at this hospital. The ICH Score is a clinical grading scale composed of factors related to a basic neuorological examination (Glasgow Coma Scale/GCS), a baseline patient characteristic (age), and initial neuroimaging (ICH volume, intraventicular hemorrhage (IVH), infratentorial or supratentorial origin). Score documentation may range from 0 to 6. The purpose of this grading scale is to provide a standard assessment tool that can be easily and rapidly determined at the time of ICH presentation by physicians without special training in stroke neurology and that will allow consistency in communication and treatment selection in clinical care and clinical research. Allowable Values: Yes or No/UTD.

• Initial ICH Score Time - The time (military time) for which the ICH score was first documented at the hospital.

Patients are eligible for the numerator population when the ICD-10-PCS Principal or Other Procedure Date and ICD-10-PCS Principal or Other Procedure Time minus the Initial Hunt and Hess Scale or ICH Score Date and Initial Hunt and Hess Scale or ICH Score Time are greater than or equal to zero minutes, OR the Initial Hunt and Hess Scale or ICH Score Date and Initial Hunt and Hess Scale or ICH Score Time minus the Arrival Date and Arrival Time are greater than or equal to zero minutes and less than or equal to 360 minutes.

S.7. Denominator Statement *(Brief, narrative description of the target population being measured)*
SAH and ICH stroke patients who arrive at this hospital emergency department (ED).

S.8. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Included Populations:
• Discharges with ICD-10-CM Principal Diagnosis Code for hemorrhagic stroke as defined in Appendix A, Table 8.2 (i.e., Table 8.2a and Table 8.2b) with or without aneurysm repair procedure (ICD-10-PCS Principal or Other Procedure Code as defined in Appendix A, Table 8.2d) or surgical intervention (ICD-10-PCS Principal or Other Procedure Code as defined in Appendix A, Table 8.2e)

11 data elements are used to calculate the denominator. Data elements and definitions:

• Admission Date: The month, day, and year of admission to acute inpatient care.

• Birthdate: The month, day, and year the patient was born.

• Comfort Measures Only: Comfort Measures Only refers to medical treatment of a dying person where the natural dying process is permitted to occur while assuring maximum comfort. It includes attention to the psychological and spiritual needs of the patient and support for both the dying patient and the patient´s family. Comfort Measures Only is commonly referred to as "comfort care" by the general public. It is not equivalent to a physician order to withhold emergency resuscitative measures such as Do Not Resuscitate (DNR).
Allowable Values: 1 (Day 0 or Day 1); 2 (Day 2 or after); 3 (Timing unclear); 4 Not documented/UTD.

• Direct Admission – Documentation that the patient was transferred from another acute care facility and taken directly to the operating room or interventional suite prior to hospital admission, or admitted directly to intensive care or other unit of the hospital. Allowable Values: Yes or No/UTD.

• Discharge Date - The month day and year the patient was discharged from acute care, left against medical advice or expired during the stay.

• Discharge Time – The documented time (military time) the patient was discharged from acute care, left against medical advice or expired during the stay.

• ED Patient - Documentation that the patient received care in a dedicated emergency department of the facility.
Allowable Values: Yes or No/UTD.

• ICD-10-CM Other Diagnosis Code: The other or secondary International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) diagnosis codes associated with the diagnosis for this hospitalization.

• ICD-10-PCS Other Procedure Codes: The International Classification of Diseases, Tenth Revision, Master Code Table (ICD-10-PCS) codes identifying all significant procedures other than the principal procedure.

• ICD-10-CM Principal Diagnosis Code: The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.

• ICD-10-PCS Principal Procedure Code: The International Classification of Diseases, Tenth Revision, Master Code Table (ICD-10-PCS) code that identifies the principal procedure performed during this hospitalization. The principal procedure is the procedure performed for definitive treatment rather than diagnostic or exploratory purposes, or which is necessary to take care of a complication.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
• Patients less than 18 years of age
• Patients who have a Length of Stay greater than 120 days
• Patients with Comfort Measures Only documented on the day of or day after hospital arrival
• Non-surgical patients discharged within 6 hours of arrival at this hospital
• Patients with admitting diagnosis of traumatic brain injury (TBI), unruptured arteriovenous malformation (AVM), and non-traumatic subdural hematoma (ICD-9-CM Other Diagnosis Codes as defined in Appendix A, Table 8.2f)

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
• Patients less than 18 years of age.
o Patient age (in years) equals Admission Date minus Birthdate.

• Patients who have a Length of Stay greater than 120 days.
o Length of Stay (in days) equals Discharge Date minus Admission Date.

• Patients with Comfort Measures Only documented:
Physician/APN/PA documentation of comfort measures only (hospice, comfort care, etc.) when the earliest day of documented CMO was on the day of arrival (Day 0) or Day after arrival (Day 1).

• Non-surgical patients discharged within 6 hours of arrival at this hospital.
o Within 6 hours of hospital arrival equals Discharge Date and Discharge Time minus Arrival Date and Arrival Time for patients who do not have an ICD-10-PCS Principal or Other Procedure Code as defined in Appendix A, Table 8.2d Aneurysm Repair Procedures or Table 8.1e Surgical Intervention Procedures.

• Patients with admitting diagnosis of traumatic brain injury (TBI), unruptured arteriovenous malformation (AVM), and non-traumatic subdural hematoma (ICD-10-CM Other Diagnosis Codes as defined in Appendix A, Table 8.2f)

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
The CSTK-03 measure is reported as an overall rate which includes SAH and ICH stoke patients for whom a severity measurement is performed prior to surgical intervention (e.g., clipping, coiling, or any surgical intervention) in patients undergoing surgical intervention and documented in the medical record; OR, documented within 6 hours of arrival at the hospital emergency department in patients who do not undergo surgical intervention. CSTK-03a and CSTK-03b are submeasures of the overall rate

measure, and stratified by the type of stroke patient as defined by the ICD-10-CM Principal Diagnosis Code in Appendix A, Table 8.2 (i.e., Table 8.2a and Table 8.2b).

Stratification Table

Measure Set #	Stratified by Principal Diagnosis Code*	
CSTK-03	Severity Measurement Performed for SAH and ICH Patients (Overall Rate)**	
CSTK-03a	Hunt and Hess Scale Performed for SAH Patients	Table 8.2a
CSTK-03b	ICH Score Performed for ICH Patients	Table 8.2b

- * Refers to the data element ICD-10-CM Principal Diagnosis Code. Each case will be stratified according to the principal diagnosis code, after the Category Assignments are completed and the overall rate calculated.
  - o ICD-10-CM Principal Diagnosis Code: The International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) code associated with the diagnosis established after study to be chiefly responsible for occasioning the admission of the patient for this hospitalization.
- ** No allowable value exists for the overall rate. It includes all diagnoses on Tables 8.2a and 8.2b.

Data element and definition:
- Initial Hunt and Hess Scale Performed - Documentation of the first Hunt and Hess Scale that was done at this hospital. The Hunt and Hess Scale is a grading system used to classify the severity of a subarachnoid hemorrhage based on the patient's clinical condition. The scale ranges from a score of 1 to 5. It is used as a predictor of prognosis and outcome with a higher grade correlating to a lower survival rate.

Data element and definition:
- Initial ICH Score Performed - Documentation of the first ICH Score that was done at this hospital. The ICH Score is a clinical grading scale composed of factors related to a basic neuorological examination (Glasgow Coma Scale/GCS), a baseline patient characteristic (age), and initial neuroimaging (ICH volume, intraventicular hemorrhage (IVH), infratentorial or supratentorial origin). Score documentation may range from 0 to 6. The purpose of this grading scale is to provide a standard assessment tool that can be easily and rapidly determined at the time of ICH presentation by physicians without special training in stroke neurology and that will allow consistency in communication and treatment selection in clinical care and clinical research.

S.13. Risk Adjustment Type (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification
If other:

S.14. Identify the statistical risk model method and variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Not Applicable

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)
Not Applicable

S.16. Type of score:
Rate/proportion
If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. Calculation Algorithm/Measure Logic *(Describe the calculation of the measure score as an ordered sequence of steps including*

identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)

Comprehensive Stroke (CSTK) Initial Patient Population Algorithm

Variable Key: Patient Age, Initial Patient Population Reject Case Flag, Length of Stay, Sub-Population 1 Flag, Sub-Population 2 Flag, and Sub-Population 3 Flag.

1. Start CSTK Initial Patient Population logic sub-routine. Process all cases that have successfully reached the point in the Transmission Data Processing Flow: Clinical which calls this Initial Patient Population Algorithm. Do not process cases that have been rejected before this point in the Transmission Data Processing Flow: Clinical.

2. Check ICD-10-CM Principal Diagnosis Code
a. If ICD-10-CM Principal Diagnosis Code is not on Table 8.1 and 8.2, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If ICD-10-CM Principal Diagnosis Code is on Table 8.1 or 8.2, continue processing and proceed to the Patient Age calculation.

3. Calculate Patient Age. Patient Age, in years, is equal to the Admission Date minus the Birthdate. Use the month and day portion of admission date and birthdate to yield the most accurate age.

4. Check Patient Age
a. If Patient Age is less than 18 years, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If Patient Age is greater than or equal to 18 years, continue processing and proceed to Length of Stay Calculation.

5. Calculate the Length of Stay. Length of Stay, in days, is equal to the Discharge Date minus the Admission Date.

6. Check Length of Stay
a. If Length of Stay is greater than 120 days, the patient is not in the CSTK Initial Patient Population and is not eligible to be sampled for the CSTK measure set. Set the Initial Patient Population Reject Case Flag to equal Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If Length of Stay is less than or equal to 120 days, the patient is in the CSTK Initial Patient Population.

7. Set the Initial Patient Population Reject Case Flag to equal No. Continue processing and proceed to the ICD-10-CM Principal Diagnosis Code to determine the CSTK sub-population.

8. Initialize Sub-Population 1 Flag, Sub-Population 2 Flag and Sub-Population 3 Flag to No.

9. Check ICD-10-CM Principal Diagnosis Code
a. If ICD-10-CM Principal Diagnosis Code is on 8.2, the patient is in the CSTK Sub-Population 3 and is eligible to be sampled for the CSTK Sub-Population 3. Set Sub-Population 3 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
b. If ICD-10-CM Principal Diagnosis Code is on 8.1, continue processing and proceed to ICD-10-PCS Principal or Other Procedure Codes.
i. If at least one ICD-10-PCS Principal or Other Procedure Codes is on Table 8.1a or 8.1b, the patient is in the CSTK Sub-Population 2 and is eligible to be sampled for the CSTK Sub-Population 2. Set Sub-Population 2 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.
ii. If none of the ICD-10-PCS Principal or Other Procedure Codes are on Table 8.1a or 8.1b, the patient is in the CSTK Sub-Population 1 and is eligible to be sampled for the CSTK Sub-Population 1. Set Sub-Population 1 Flag to Yes. Return to Transmission Data Processing Flow: Clinical in the Data Transmission section.

CSTK-03: Severity Measurement Performed for SAH and ICH Patients (Overall Rate)

Numerator: The number of SAH and ICH stroke patients for whom a severity measurement is performed prior to surgical intervention in patients undergoing surgical intervention and documented in the medical record, OR documented within 6 hours of hospital arrival for patients who do not undergo surgical intervention.

Denominator: SAH and ICH stroke patients who arrive at this hospital emergency department (ED)

Variable Key: Timing I, Timing II, Timing III, Timing IV, Timing V

1. Start processing. Run cases that are included in the Comprehensive Stroke (CSTK) Initial Patient Population and pass the edits defined in the Transmission Data Processing Flow: Clinical through this measure.

2. Check ICD-10-CM Principal Diagnosis Code
a. If ICD-10-CM Principal Diagnosis Code is not on Table 8.2, the case will proceed to a Measure Category Assignment of B for Overall Rate CSTK-03 and will not be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If ICD-10-CM Principal Diagnosis Code is on Table 8.2, continue processing and proceed to ICD-10-PCS Other Diagnosis Code.

3. Check ICD-10-PCS Other Diagnosis Code
a. If ICD-10-PCS Other Diagnosis Code is on Table 8.2f, the case will proceed to a Measure Category Assignment of B for Overall Rate CSTK-03 and will not be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If ICD-10-PCS Other Diagnosis Code is not on Table 8.2f or all missing, continue processing and proceed to ED Patient.

4. Check ED patient
a. If ED Patient is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If ED Patient equals No, continue processing and proceed to Step 5 to check direct admission.
c. If ED Patient equals Yes, continue processing and proceed to Step 6 to check Comfort Measures Only.

5. Check Direct Admission
a. If Direct Admission is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Direct Admission equals No, the case will proceed to a Measure Category Assignment of B for Overall Rate CSTK-03 and will not be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Direct Admission equals Yes, continue processing and proceed to Comfort Measures Only.

6. Check Comfort Measures Only
a. If Comfort Measures Only is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Comfort Measures Only equals 1, the case will proceed to a Measure Category Assignment of B for Overall Rate CSTK-03 and will not be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Comfort Measures Only equals 2, 3, or 4, continue processing and proceed to check ICD-10-CM Principal or Other Procedure Codes.

7. Check ICD-10-PCS Principal or Other Procedure Codes
a. If all missing or none ICD-10-PCS Principal or Other Procedure Codes is on Table 8.2d or 8.2e, continue processing and proceed to Step 20 to check Discharge Date.
b. If any ICD-10-PCS Principal or Other Procedure Codes is on Table 8.2d or 8.2e, continue processing and proceed to Initial Hunt and Hess Scale Performed.

8. Check Initial Hunt and Hess Scale Performed
a. If Initial Hunt and Hess Scale Performed is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Initial Hunt and Hess Scale Performed equals No, continue processing and proceed to step 14 to check Initial ICH Score

Performed.
c. If Initial Hunt and Hess Scale Performed equals Yes, continue processing and proceed to check Initial Hunt and Hess Scale Date.

9. Check Initial Hunt and Hess Scale Date
a. If Initial Hunt and Hess Scale Date is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Initial Hunt and Hess Scale Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Initial Hunt and Hess Scale Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to check Initial Hunt and Hess Scale Time.

10. Check Initial Hunt and Hess Scale Time
a. If Initial Hunt and Hess Scale Time is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Initial Hunt and Hess Scale Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Initial Hunt and Hess Scale Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to check ICD-10-PCS Principal or Other Procedure Code Date.

11. Check ICD-10-PCS Principal or Other Procedure Code Date
a. If ICD-10-PCS Principal or Other Procedure Code Date is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If ICD-10-PCS Principal or Other Procedure Code Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If ICD-10-PCS Principal or Other Procedure Code Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to check ICD-10-PCS Principal or Other Procedure Code Time.

12. Check ICD-10-PCS Principal or Other Procedure Code Time
a. If ICD-10-PCS Principal or Other Procedure Code Time is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If ICD-10-PCS Principal or Other Procedure Code Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If ICD-10-PCS Principal or Other Procedure Code Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to the Timing I calculation.

13. Calculate Timing I. Timing I, in minutes, is equal to the ICD-10-PCS Principal or Other Procedure Code Date and the ICD-10-PCS Principal or Other Procedure Code Time minus the Initial Hunt and Hess Scale Date and Initial Hunt and Hess Scale Time.
a. If the time in minutes is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If the time in minutes is less than zero, the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If the time in minutes is greater than or equal to zero, the case will proceed to a Measure Category Assignment of E for Overall Rate CSTK-03 and will be in the numerator population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.

14. Check Initial ICH Score Performed
a. If Initial ICH Score Performed is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03

and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Initial ICH Score Performed equals No, the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Initial ICH Score Performed equals Yes, continue processing and proceed to check Initial ICH Score Date.

15. Check Initial ICH Score Date
a. If Initial ICH Score Date is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Initial ICH Score Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Initial ICH Score Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to check Initial ICH Score Time.

16. Check Initial ICH Score Time
a. If Initial ICH Score Time is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Initial ICH Score Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Initial ICH Score Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to check ICD-10-PCS Principal or Other Procedure Code Date.

17. Check ICD-10-PCS Principal or Other Procedure Code Date
a. If ICD-10-PCS Principal or Other Procedure Code Date is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If ICD-10-PCS Principal or Other Procedure Code Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If ICD-10-PCS Principal or Other Procedure Code Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to check ICD-10-PCS Principal or Other Procedure Code Time.

18. Check ICD-10-PCS Principal or Other Procedure Code Time
a. If ICD-10-PCS Principal or Other Procedure Code Time is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If ICD-10-PCS Principal or Other Procedure Code Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If ICD-10-PCS Principal or Other Procedure Code Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to the Timing II calculation.

19. Calculate Timing II. Timing II, in minutes, is equal to the ICD-10-PCS Principal or Other Procedure Code Date and the ICD-10-PCS Principal or Other Procedure Code Time minus the Initial ICH Score Date and Initial ICH Score Time.
a. If the time in minutes is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If the time in minutes is less than zero, the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If the time in minutes is greater than or equal to zero, the case will proceed to a Measure Category Assignment of E for overall rate CSTK-03 and will be in the numerator population. Continue processing and proceed to Step 33 to Initialize Measure Category

Assignment for strata measure CSTK-03a and CSTK-03b.

20. Check Discharge Date
a. If Discharge Date is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Discharge Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Discharge Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to check Discharge Time.

21. Check Discharge Time
a. If Discharge Time is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Discharge Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Discharge Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to check Arrival Date.

Continued in Section Ad.8 Additional Information/Comments.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available at measure-specific web page URL identified in S.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals that choose to sample have the option of sampling quarterly or sampling monthly. A hospital may choose to use a larger sample size than is required. Hospitals whose Initial Patient Population size is less than the minimum number of cases per quarter for the measure set cannot sample. Regardless of the option used, hospital samples must be monitored to ensure that sampling procedures consistently produce statistically valid and useful data. Due to exclusions, hospitals selecting sample cases MUST submit AT LEAST the minimum required sample size.

The sub-population for the CSTK-03 measure Initial Patient Population is CSTK 3-Hemorrhagic Stroke. The CSTK 3 sub-population sampling group includes patients admitted to the hospital for inpatient acute care if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.2, a Patient Age (Admission Date – Birthdate) > 18 years and a Length of Stay (Discharge Date - Admission Date) less than or equal to 120 days.

Quarterly Sampling
The Quarterly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for CSTK sub-population 3. Hospitals performing quarterly sampling for the CSTK-03 measure must ensure that its Initial Patient Population and sample size meet the following conditions for the CSTK 3 sub-population sampling group:

Quarterly Sample Size Based on CSTK Sub-population 3 for Hemorrhagic Stroke (Table 3)

Sub-Population 3: If "N" > 750, then 'n' 150
Minimum Required Sample Size: 150 records

Sub-Population 3: If "N" 376-750, then 'n' 20%
Minimum Required Sample Size: 20% of Sub-Population records

Sub-Population 3: If "N" 76-375, then 'n' 75
Minimum Required Sample Size: 75 records

Sub-Population 3: If "N" < 75, then 'n' 100%

Minimum Required Sample Size: No sampling; 100% Sub-Population records required

Monthly Sampling
The Monthly Sample Size "n", i.e., Minimum Required Sample Size, is based on the Initial Patient Population Size "N" for CSTK sub-population 3. Hospitals performing monthly sampling for the CSTK-03 measure must ensure that its Initial Patient Population and sample size meet the following conditions for each sub-population sampling group:

Monthly Sample Size Based on CSTK Sub-population 3 for Ischemic Stroke (Table 6)

Sub-Population 3: If "N" > 250, then 'n' 50
Minimum Required Sample Size: 50 records

Sub-Population 3: If "N" 126-250, then 'n' 20%
Minimum Required Sample Size: 20% of Sub-Population size records

Sub-Population 3: If "N" 26-125, then 'n' 25
Minimum Required Sample Size: 25 records

Sub-Population 3: If "N" < 25, then 'n' 100%
Minimum Required Sample Size: No sampling; 100% Sub-Population records required

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not Applicable

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Any file with missing data will result in a measure category assignment of X and rejection of the file from the warehouse, unless the data are not used to process the measure. If the data are used to process the measure and have been reported by the abstractor as No/UTD, the case will result in a measure category assignment of D (i.e., failed measure, not rejected).

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Paper Medical Records

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
A web-based data collection tool was developed by The Joint Commission for the pilot test process. Currently, hospitals have the flexibility of creating their own tool modeled after the pilot tool or they may develop their own data collection tools using the data element dictionary and allowable values specified in the implementation guide. Hospitals also have the option of selecting a vendor-developed data collection tool which has been verified by The Joint Commission.

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility, Population : National

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules,*

or calculation of individual performance measures if not individually endorsed.)
Not Applicable

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
2866_MeasureTesting_MSF6.5.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b6)

Measure Title: CSTK-03: Severity Measurement Performed for SAH and ICH Patients (Overall Rate)
Date of Submission: 1/15/2016
Type of Measure: Process

☐ Composite – ***STOP – use composite testing form***	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. Risk models should not obscure disparities in care for populations by including factors that are associated with differences/inequalities in care, such as race, socioeconomic status, or gender (e.g., poorer treatment outcomes of African American men with prostate cancer or inequalities in treatment for CVD risk factors between men and women). It is preferable to stratify measures by race and socioeconomic status rather than to adjust out the differences.

16. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation

counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Not Applicable

1.3. What are the dates of the data used in testing? October 1, 2012 –March 31, 2013; first and second quarter 2015.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Hospitals were recruited for the pilot test of the measures via an open call on The Joint Commission web site. An announcement of the call with a link to The Joint Commission web site was also posted on the American Heart Association/American Stroke Association web site. Hospitals were selected with the intent to capture variability related to ownership, size, type of facility and location. Eighty-two hospitals from twenty-seven states were selected from more than 120 volunteers to participate in the six-month pilot test of the measures. Twenty hospitals withdrew during the pilot test citing lack of resources to complete the project. Sixty-two hospitals submitted data for each month of the six-month pilot test. An additional four hospitals submitted data for one or more months.

Sixty-six hospitals contributed data for the analysis of the measures:

Ownership:	
For Profit	14
Not for Profit	52

Bed Size:	
Less than 100	0
100 – 199	3
200 – 299	7
300 – 499	26
Greater than 500+	30

Located in 27 states:
Alabama
Arizona
California
Colorado
Florida
Georgia
Illinois
Indiana
Kentucky
Louisiana
Maryland
Massachusetts
Michigan
Minnesota
Missouri
Nevada
New Jersey
New York
North Carolina
Ohio
Oregon
Pennsylvania
Tennessee

Texas
West Virginia
Washington
Wisconsin

Other:

Teaching	41
Non-teaching	25
Urban	59
Rural	7

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

During the six-month pilot test, sixty-six hospitals submitted data for 2,471 inpatient records. The cases included patients greater than 18 years of age, male and female, all races, and all payers. Age, gender, racial, and payer distribution are not known because the results were de-identified. The patient population was comprised of all patients discharged with an ICD-9-CM Principal Diagnosis Code for subarachnoid hemorrhage or intracerebral hemorrhage.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Data from the six month pilot test of the measures were used for reliability testing and face validity. To test the empirical validity of the measures, two quarters of data from 42 Joint Commission certified comprehensive stroke centers were used to conduct a secondary analysis.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Reliability testing was performed at twelve participating pilot sites. Testing was conducted on a stratified random sample of records selected from each organization at the organization and measure category level. Hospitals were visited by teams of two Joint Commission staff during April, May, June, July, and August 2013. During these visits, Joint Commission staff re-abstracted records previously abstracted by hospital staff. A total of 281 records were re-abstracted. In cases of disagreement between the re-abstracted and original abstraction on a data element, reasons for disagreement were determined and adjudication was made as to whether original

or re-abstraction findings were correct. Reliability was addressed by comparing the original abstracted and adjudicated re-abstracted values, with the adjudicated value serving as the gold standard. The data analysis included both the percent agreement and the kappa statistic to adjust for chance agreement for categorical data elements.

2a2.3. For each level checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data Element	Number of Mismatches	Match Rate	Kappa
Arrival Date	7	97.5%	NA
Arrival Time	50	82.2%	NA
ICD-9-CM Other Procedure Dates	15	94.1%	NA
ICD-9-CM Other Procedure Times	39	86.0%	NA
ICD-9-CM Principal Procedure Date	15	94.1%	NA
ICD-9-CM Principal Procedure Time	15	94.1%	NA
Initial Hunt and Hess Scale Date	12	95.7%	NA
Initial Hunt and Hess Scale Performed	2	97.9%	0.91
Initial Hunt and Hess Scale Time	21	92.5%	NA
Initial ICH Score Date	12	95.7%	NA
Initial ICH Score Performed	8	97.2%	0.86
Initial ICH Score Time	21	94.3%	NA
Discharge Date	9	96.8%	NA
Discharge Time	80	71.5%	NA
ED Patient	2	99.3%	0.96

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

A statistical measure of inter-rater reliability is Cohen's Kappa, which ranges generally from 0.0 to 1.0 (although negative numbers are possible), where large values mean better reliability and values near zero suggest that agreement is attributable to chance alone. It indicates the proportion of agreement not expected by chance alone (e.g., Kappa of 0.6 means that raters agreed 60% of the time over and above what would be expected by chance alone).

Landis & Koch, 1977 offers the following clarification of Kappa interpretation:

< 0	Poor agreement
0.00-0.20	Slight agreement
0.21-0.40	Fair agreement
0.41-0.60	Moderate agreement
0.61-0.80	Substantial agreement
0.81-1.00	Almost perfect agreement

Other authors (Cicchetti & Sparrow; Fleiss) have suggested additional classifications for interpreting the Kappa statistic, but all seem to indicate Kappa > 0.60 is desirable.

The statistical measure of inter-rater reliability is suitable for the categorized data type, therefore the measure is applied to the following data elements: *Initial Hunt and Hess Scale Performed, Initial ICH Score Performed,* and *ED Patient*. The Kappa values indicate almost perfect agreement for these 3 data elements. According to the statistical measure of Kappa value, we believe these results demonstrate acceptable reliability of the assessment data used in the performance measure.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

X **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

X **Empirical validity testing**

X **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Face Validity:

Measure face validity was assessed via survey and focus groups of hospitals participating in the pilot test. Focus group discussions were held at all test sites visited, during which we received feedback as to whether the measure, data elements, and definitions accurately reflected existing evidence. All of the respondents indicated that all aspects of the measures accurately reflected current evidence. Measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant.

To determine feasibility and identify areas for potential revision, test sites were asked to electronically rate the clarity of numerator statements, denominator statements, and measure information forms (MIFs) on a five point Likert scale (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree). Data elements and associated tables were evaluated for clarity, accuracy, data availability and accessibility.

Empirical Validity:

Measure convergent validity was assessed using hospitals patient level data from The Joint Commission warehouse, measure specifications, including population identification, numerator and denominator statements, exclusions, and data elements and their definitions were found to be understandable, retrievable, and relevant in previous face validity testing. We conducted a secondary analysis to help interpret results; correlation among CSTK process measures.

The data comprise first and second quarter 2015. 42 hospitals submitting 65,389 inpatient records for the selected CSTK measures CSTK-1, CSTK-2, CSTK-3 and CSTK-6. The hospital's selection was based on those hospitals that reported 6 months of data and had 30 or more denominator cases for the period.

Comprehensive Stroke (CSTK) Initial Patient Population

The CSTK Initial Patient Population is unique in that it is comprised of three distinct subpopulations: ischemic stroke patients who do not undergo a reperfusion therapy (i.e., procedure), ischemic stroke patients who

undergo a reperfusion therapy (IV t-PA, IA t-PA, or mechanical endovascular reperfusion (MER) therapy), and hemorrhagic stroke patients.

Subpopulation 1: Ischemic Stroke Without Procedure
This subpopulation comprises ischemic stroke patients who are admitted to the hospital for inpatient acute care and do not undergo a reperfusion procedure (CSTK-01 measure). Patients are included in the CSTK-1 Ischemic Stroke Without Procedure subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.1, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

Subpopulation 2: Ischemic Stroke With IV t-PA, IA t-PA, or MER
This subpopulation comprises ischemic stroke patients who receive IV t-PA, IA t-PA, or MER procedures during the hospital stay (CSTK-01 and CSTK-02 measures). Patients admitted to the hospital for inpatient acute care are included in the CSTK-2 Ischemic Stroke With IV t-PA, IA t-PA, or MER subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.1 and ICD-10-PCS Principal or Other Procedure Codes as defined in Appendix A, Table 8.1a OR Table 8.1b, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

Subpopulation 3: Hemorrhagic Stroke
This subpopulation comprises hemorrhagic stroke patients admitted to the hospital for inpatient acute care (CSTK-03 and CSTK-06 measures). Patients are included in the CSTK-3 -Hemorrhagic Stroke subpopulation sampling group if they have: ICD-10-CM Principal Diagnosis Code as defined in Appendix A, Table 8.2, a Patient Age (Admission Date – Birthdate) ≥ 18 years and a Length of Stay (Discharge Date - Admission Date) ≤ 120 days.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)
Pilot Test Data:
Descriptive statistics for the measure: N= 66 hospitals
Overall rate=20% (SD=23%), min=0%, max=100%

Additional analyses:
CSTK03 and CSTK06 are measures for hemorrhagic stroke. The measure rate among these measures are expected to be correlated. The Pearson correlation coefficient is calculated.

Pearson Correlation Coefficient interpretation: the range is from -1 to 1. When the correlation coefficient is close to +1 or -1, it means that there is strong correlation; p value is utilized to determine if the corrlation coefficient is significant or not. If it is less than 0.05, then the conclusion usually is significant.

The table below demonstrates that CSTK03 rate was negtively and slightly correlated to rest of these measures. Because the data is not normally distributed, the corrlation coefficient on correlation may not be the best measure of association. Also the large proportion of zero measure rates for some of the measures, while demonstrating considereable room for improvement, would tend to skew the correlations.

Pearson Correlation Coefficients, N = 66 Prob > |r| under H0: Rho=0

	CSTK03	**CSTK06**
Correlation Coefficient	1.00000	-0.02229

P value		0.8590

The average rating for measure CSTK-03 numerator and denominator statements, including the clarity of numerator and denominator inclusions and exclusions was 3.74. Although this rating is slightly lower than those for other measures in the CSTK set, it can still be used to distinguish good and poor quality.

Data element validity was evaluated for clarity and accuracy, as well as, data availability and accessibility. The percentage of agreement for new data elements developed specifically for CSTK-03 and other measures in the CSTK measure set is detailed in the table below:

Data Element Name	Clarity / Accuracy	Availability / Accessibility
Discharge Time	95.45%	37.21%
ICD-9-CM Other Procedure Times	90.91%	37.21%
ICD-9-CM Principal Procedure Time	100.00%	37.21%
Initial Hunt and Hess Scale Date	95.45%	37.21%
Initial Hunt and Hess Scale Performed	97.73%	37.21%
Initial Hunt and Hess Scale Time	90.91%	37.21%
Initial ICH Score Date	97.67%	37.21%
Initial ICH Score Performed	95.35%	37.21%
Initial ICH Score Time	95.24%	37.21%

1Q and 2Q 2015 Data:
Overall descriptive statistics for CSTK selected measures: N = 42 certified comprehensive stroke hospitals: n = 65,389

CSTK-01
Median: 89%
Percentile 10%: 70%
Percentile 25%: 80%
Percentile 75%: 93%
Percentile 90%: 95%

CSTK-02
Median: 95%
Percentile 10%: 63%
Percentile 25%: 75%
Percentile 75%: 100%
Percentile 90%: 100%

CSTK-03
Median: 61%
Percentile 10%: 33%

Percentile 25%: 48%
Percentile 75%: 79%
Percentile 90%: 89%

CSTK-06
Median: 86%
Percentile 10%: 74%
Percentile 25%: 80%
Percentile 75%: 93%
Percentile 90%: 100%

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
CSTK_1	42	0.83687	0.17097	35.14857	0.00267	1.00000
CSTK_2	41	0.85894	0.20516	35.21648	0.20000	1.00000
CSTK_3	42	0.59308	0.22979	24.90933	0.07273	0.92969
CSTK_6	42	0.83555	0.17123	35.09314	0	1.00000

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	CSTK_1	CSTK_2	CSTK_3	CSTK_6
CSTK_1	1.00000 42	0.41205 0.0074 41	0.65340 <.0001 42	0.69186 <.0001 42
CSTK_2	0.41205 0.0074 41	1.00000 41	0.28658 0.0693 41	0.31724 0.0433 41
CSTK_3	0.65340 <.0001 42	0.28658 0.0693 41	1.00000 42	0.55910 0.0001 42
CSTK_6	0.69186 <.0001 42	0.31724 0.0433 41	0.55910 0.0001 42	1.00000 42

The CSTK measures table shows a positive correlation and statistical significance which indicates that hospitals with high quality on one CSTK measure tend to have high correlations on the other stroke measures.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)
Pilot Test Data:
We note that ratings for both data elements and the measure are relatively high; numerators, denominators and measures were ranked above the midpoint of 3.0 (average), and data elements were above 75% positive in

clarity, collectability, and correctness of data sources. We conclude that the measures and specifications are valid.

The rating for data availability and accessibility reflects that the ICH Score was not always documented in the medical record; however, documentation of the Hunt and Hess Scale was consistent and not an issue. Additionally, *ICD-9-CM Other Procedure Times* were often missing, although the *ICD-9-CM Principal Procedure Time* was consistently documented. We expect that data availability will improve with use over time.

1Q and 2Q 2015 Data:
Overall the positive inter-correlations indicates convergent validity of all the measures.
They are positively correlated with other evidence-based processes of care.

2b3. EXCLUSIONS ANALYSIS .
NA □ no exclusions — *skip to section 2b4*
2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Pilot Test Data:
The following exclusions were analyzed for frequency and variability across providers:

- Patients less than 18 years of age
- Patients who have a Length of Stay greater than 120 days
- Non-surgical patients discharged within 6 hours of arrival at this hospital
- Patients with admitting diagnosis of traumatic brain injury (TBI), unruptured arteriovenous malformation (AVM), and non-traumatic subdural hematoma (ICD-9-CM Other Diagnosis Codes as defined in Appendix A, Table 8.2f)

1Q and 2Q 2015 Data:
The following exclusions were analyzed by subpopulation and measure for frequency and variability across providers:

- Patients less than 18 years of age
- Patients who have a Length of Stay greater than or equal to 120 days
- Patients with *Comfort Measures Only* documented on the day of or day after hospital arrival
- Non-surgical patients discharged within 6 hours of arrival at this hospital
- Patients with admitting diagnosis of traumatic brain injury (TBI), unruptured arteriovenous malformation (AVM),and non-traumatic subdural hematoma (ICD-9-CM Other Diagnosis Codes as defined in Appendix A, Table 8.2f)

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Pilot Test Data:
There were 10668 admissions included in the initial cohort. From among the 10668 admissions in 66 hospitals, the descriptive statistics are given below.

Exclusion: Patient not in CSTK Initial Patient Population

Note: A case was excluded from the Initial Patient Population as determined by the following:

- Patients less than 18 years of age: Overall Occurrence n = 39 (0.37%)
- Patients who have a Length of Stay greater than 120 days: Overall Occurrence n = 12 (0.11%)

There were 3830 admissions included in the initial cohort and diagnosed hemorrhagic stroke. From among the 3830 admissions in 66 hospitals, the descriptive statistics are given below.

Exclusion: Non-surgical patients discharged within 6 hours of arrival at this hospital
Overall Occurrence n = 9
Overall Occurrence Percentage: 0.23%

Exclusion: Patients with admitting diagnosis of traumatic brain injury (TBI), unruptured arteriovenous malformation (AVM), and non-traumatic subdural hematoma
Overall Occurrence n = 15
Overall Occurrence Percentage: 0.39%

1Q and 2Q 2015 Data:
There were 65,389 admissions selected from the initial cohort. From among the 65,389 admissions in 42 hospitals, the descriptive statistics are given below.

Applied To Measure CSTK-01 CSTK-03 CSTK-06
Exclusion: Comfort Measures - 1 Day 0 or 1
Overall Occurrence n = 1,300
Overall Occurrence Percentage: 2%
Minimum: 0.47%
Median: 3%
Maximum: 6%

Applied To Measure CSTK-3
Exclusion: Non-surgical patients discharged within 6 hours of arrival at this hospital
Overall Occurrence n = 20,266
Overall Occurrence Percentage: 62%
Minimum: 47%
Median: 54%
Maximum: 17.39%
NMISS= 20,232 missing observation

Applied To Measure CSTK-3
Exclusion: Patients with admitting diagnosis of traumatic brain injury (TBI), unruptured arteriovenous malformation (AVM), and non-traumatic subdural hematoma (ICD-9-CM Other Diagnosis Codes as defined in Appendix A, Table 8.2f)
None

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Pilot Test Data:
According to overall occurrences in 2b3.2, the overall frequency of exclusions are low for those in the measure denominator. The distribution of exclusions across hospitals is very narrow indicating that the occurrence is random and likely would not bias performance results.

It is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a Length of Stay greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Non-surgical patients discharged within 6 hours of arrival at this hospital
Rationale: It is inappropriate to include patients who do not undergo surgical intervention whose hospital stay is less than the timeframe for performing severity measurement.

Patients with admitting diagnosis of traumatic brain injury (TBI), unruptured arteriovenous malformation (AVM), and non-traumatic subdural hematoma
Rationale: It is inappropriate to include patients who do present to the hospital with brain hemorrhage due to trauma.

1Q and 2Q 2015 Data:
The overall frequency of exclusions is low for those in the measure denominator. The distribution of exclusions across hospitals is narrow indicating that the occurrence is random and likely would not bias performance results.

It is believed that all of the exclusions should be retained for the following reasons:

Patients less than 18 years of age
Rationale: The literature does not support use in the pediatric patient.

Patients who have a *Length of Stay* greater than 120 days
Rationale: Included for this measure in order to harmonize with other CMS/Joint Commission aligned measures.

Patients with *Comfort Measures Only* documented
Rationale: It is inappropriate to treat such patients beyond those measures necessary for comfort.

Non-surgical patients discharged within 6 hours of arrival at this hospital
Rationale: Non-surgical patients are excluded when the duration of stay is less than the timeframe specified for obtaining and documenting the initial Hunt and Hess Scale or ICH Score.

Patients with admitting diagnosis of traumatic brain injury (TBI), unruptured arteriovenous malformation (AVM), and non-traumatic subdural hematoma (NTSDH) (ICD-9-CM Other Diagnosis Codes as defined in Appendix A, Table 8.2f)
Rationale: Patients with TBI, unruptured AVM, and NTSDH are excluded to prevent false inclusion of cases when brain hemorrhage is due to trauma, or the patient does not have an acute hemorrhage

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

Not Applicable

2b4.1. What method of controlling for differences in case mix is used?

X **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Not Applicable

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care and not related to disparities*)

Not Applicable

2b4.4. What were the statistical results of the analyses used to select risk factors?

Not Applicable

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Not Applicable

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

if stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not Applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not Applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not Applicable

2b4.9. Results of Risk Stratification Analysis:

Not Applicable

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.e., *what do the results mean and what are the norms for the test conducted*)
Not Applicable

***2b4.11. Optional Additional Testing for Risk Adjustment** (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods*)

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*
Descriptive statistics for the performance measure scores for all tested entities were constructed. These statistics were the mean, standard deviation, median, minimum, and maximum scores.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)
A meaningful difference was defined as a spread of more than 30 percentage points between minimum and maximum scores or between the median and maximum pilot hospital rates.

Results for CSTK-03 were:
Descriptive statistics for measure: N=66 hospitals
Overall rate=20.0%
Standard Deviation=23.4%
Minimum=0%
Median=10.6%
Maximum=100%

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.e., *what do the results mean in terms of statistical and meaningful differences?*)

The results are interpreted as showing a meaningful spread between both the minimum and maximum scores and between the median and maximum scores. This is indicative considerable room for improvement in performance.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Not Applicable

Note: *This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different datasources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not Applicable

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Not Applicable

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not Applicable

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry), Other

If other: Data element allowable values are selected, either manually or electronically, from clinical and coded data available in medical record documentation. All medical record documentation is used in the abstraction process. Vendor data collection tools are used to import data elements needed for measure rate calculation.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

Some data elements are in defined fields in electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

The Joint Commission recognizes that not all hospitals currently have the capacity to abstract this measure electronically, so offers a chart-abstracted version which allows for data capture from unstructured data fields. The Joint Commission plans to retool the measure for capture from electronic sources within the next several years.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

This measure was collected during a 6 month pilot process for Joint Commission Comprehensive Stroke Certification. It was learned via discussions and written pilot evaluation materials that there were some issues related to data abstraction for the following data elements:

- Initial Hunt and Hess Scale Performed
  - Clarification was requested regarding the use of abbreviations and alternative terms for the Hunt and Hess Scale.
- Initial ICH Score Performed
  - Clarification was requested regarding the components of the ICH score and how to calculate the ICH score.

These issues related to data abstraction have been easily resolved through clarification of guidelines for abstraction.

Based on feedback from the pilot sites, the measure algorithm was modified to add the following data elements:
• Direct Admission
o Addition of this data element was requested to capture in the denominator population those stroke patients transferred from another acute care facility and taken directly to the interventional suite or operating room, bypassing the emergency department prior to hospital admission.
• Comfort Measures Only
o Addition of this data element was requested to exclude stroke patients who are Comfort Measures Only on the day of or day after hospital arrival.

Other information impacting the feasibility and implementation of the measure was also obtained from the pilot process and is summarized as follows:

Staff Training and Education:
To prepare for and support continuous data collection throughout the pilot test, a total of ten hours were spent on staff training and education. Training was accomplished via two 2-hour webinars and monthly conference calls with pilot site participants. Additionally, some sites reported that physician education was needed to ensure Hunt and Hess Scale and ICH Score completion and documentation.

Case Identification/Medical Record Retrieval:
Case identification was not a problem; cases were identified by the ICD-9-CM principal diagnosis codes for subarachnoid hemorrhage (430) and intracerebral hemorrhage (431). Record retrieval time varied depending on the type of medical record. On average, 10 minutes were spent for record retrieval with more time spent to retrieve a paper record than electronic health record.

Case Selection:
For the pilot test of the measures, 100% record review without sampling was requested. During the pilot process it was noted that some facilities treated more than 25 hemorrhagic stroke patients per month. A sampling methodology for the hemorrhagic sub-population was added to the measure specifications post-pilot to ease abstraction burden for hospitals with a large number of hemorrhagic cases.

Data Abstraction:
Medical record review to abstract all data elements required for the measure set averaged 45 minutes per record. Time spent on record review varied with case complexity and the number of procedures and interventions performed, as well as, the number of data elements collected for the meausre. In general, hemorrhagic stroke cases required less time for review than did ischemic stroke cases.

Data abstraction was primarily done by nurses, (e.g., Registered Nurse(s) with a quality improvement background, Stroke Coordinators, and Advanced Practice Nurses). Some pilot sites reported that the abstractor reviewed the record with the medical director or neurologist at least initially to identify documentation of measure specific data elements. Data specialists or administrative staff were utilized to enter abstracted data into the on-line data collection tool.

Cost of Data Abstraction:
Using 2012 national wage averages, it is estimated that the cost per case to abstract for this measure was approximately $3.50.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.
This measure is one in a set of measures implemented with discharges effective January 1, 2015 for Joint Commission Comprehensive Stroke Certification. There are no fees or licensing requirements to use the Joint Commission performance measures, all of which are in the public domain.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Public Health/Disease Surveillance Quality Improvement with Benchmarking (external benchmarking to multiple organizations)	Quality Improvement (Internal to the specific organization) Disease-Specific Care Certification for Comprehensive Stroke Centers http://www.jointcommission.org/certification/dsc_home.aspx

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

• Name of program and sponsor: Disease-Specific Care Certification for Comprehensive Stroke Centers; The Joint Commission
• Purpose: A certification program that recognizes the specific capabilities of hospitals that treat the most complex stroke cases.
• Geographic area and number and percentage of accountable entities and patients included: Nationwide; 95 hospitals

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

Not Applicable

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Not Applicable

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

The initial performance measure gap for this measure was huge at ~80%. Although considerable improvement has been noted since the pilot test of the measure based on the first two quarters of Joint Commission ORYX performance measurement data, the

performance gap remains significant, especially for ICH patients. For 2Q2015 (N=51 hospitals), a wide range of performance was noted with a 10% gap at the 90th percentile and 70% gap at the 10th percentile.

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
Not Applicable

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
There were no unintended negative consequences reported or detected during testing or since implementation of the measure specifications.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Not applicable

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide

a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
Not applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Joint Commission
Co.2 **Point of Contact:** Karen, Kolbusz, kkolbusz@jointcommission.org, 630-792-5931-
Co.3 **Measure Developer if different from Measure Steward:** The Joint Commission
Co.4 **Point of Contact:** Ann, Watt, awatt@jointcommission.org, 630-792-5944-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The role of the Technical Advisory Panel was to provide advisory oversight in the literature review, measure construct and content, review of testing results, and endorsement of draft and finalized measures. Additionally they may be called upon in the future to provide measure content oversight and updates.

Technical Advisory Panel

Harold P. Adams, Jr., MD
Professor of Neurology
Director of the Division of Cerebrovascular Diseases
Director of the Stroke Center
University of Iowa Health Care
Iowa City, IA

Mark J. Alberts, MD
Clinical Vice-Chair
Department of Neurology and Neurotherapeutics
The University of Texas Southwestern Medical Center
Dallas, TX

Anne W. Alexandrov, PhD, RN, FAAN
Health Outcomes Institute
Fountain Hills, AZ

Andrew W. Asimos, MD
Director of Emergency Stroke Care
Carolinas Medical Center
Charlotte, NC

Thomas P. Bleck, MD, FCCM
Professor of Neurological Science, Neurosurgery, Internal Medicine, and Anesthesiology
Associate Chief Medical Officer, Critical Care
Rush Medical College
Chicago, IL

Darlene Bradley, RN, PhDc, MSN, MAOM
Director Emergency/Trauma Services
University of California Irvine Medical Center
Orange, CA

Mary G. George, MD, MSPH FACAS
Medical Officer
Division for Heart Disease and Stroke Prevention
Centers for Disease Control and Prevention
Atlanta, GA

Martin Gizzi, MD, PhD
Legacy Health System
Portland, OR

James, Grotta, MD
Professor and Chairman, Department of Neurology
Chief, Vascular Neurology
University of Texas Medical School-Houston
Houston, TX

Edward C. Jauch, MD, MS
Professor, Division of Emergency Medicine and Department of Neurosciences
Medical University of South Carolina
Charleston, SC

Irene Katzan, MD, MS
Director, Neurological Institute Center for Outcomes Research and Evaluation
Cleveland Clinic
Cleveland, OH

Jawad F. Kirmani, MD
Director of Stroke and Cerebrovascular Program
JKF Medical Center
Edison, NJ

Brajesh K. Lal, MD
University of Maryland Medical Center,
Division of Vascular Surgery
Baltimore, MD

Kimberly E. Levasseur-Franklin, PharmD, BCPS
Critical Care Pharmacist, Neuro/Spine ICU
Clinical Leader, ICU Pharmacist Work Group
Northwestern Memorial Hospital
Chicago, IL

Barbara Lutz, PhD, RN, CRRN, APHN-BC, FAHA, FNAP, FAAN
McNeill Distinguished Professor
School of Nursing / College of Health and Human Services
University of North Carolina-Wilmington
Wilmington, NC

Cameron McDougall, MD, FRCSC
Barrow Neurological Institute
Phoenix, AZ

Kathy Morrison, RN, MSN, CNRN
Stroke Program Coordinator
Penn State Hershey Medical Center
Hershey, PA

Charles J. Prestigiacomo, MD, FAANS, FACS
Professor, Department of Neurological Surgery, Radiology, Neurology and Neurosciences
Director, Residency Program in Neurological Surgery
Director, Cerebrovascular and Endovascular Surgery
University of Medicine and Dentistry of New Jersey
Newark, NJ

Howard A. Rowley, MD
Chief of Neuroradiology
Joseph Sackett Professor of Radiology
Professor of Radiology, Neurology, and Neurosurgery
University of Wisconsin
Madison, WI

Jeffrey Saver, MD, FAHA, FAAN
Professor of Neurology
Stroke Center Director
Professor of Medicine
David Geffen School of Medicine at UCLA,
Ronald Reagan UCLA Medical Center
Los Angeles, CA

Lee H. Schwamm, MD, FAHA
Vice Chair of Neurology, Professor of Neurology
Harvard Medical School
Chair, American Heart Association GWTG Steering Committee
Director, Stroke Program/Telemedicine
Massachusetts General Hospital
Boston, MA

Deborah Summers, MSN, ACNS-BC, CNRN, CCRN, FAHA
Stroke Program Coordinator
St. Luke's Brain and Stroke Institute
Kansas City, MO

Dileep R. Yavagal, MD
Director, Interventional Neurology
Assistant Professor, Clinical Neurology and Neurosurgery
University of Miami and Jackson Memorial Hospital
Miami, FL

Osama O. Zaidat, MD, MS, FAAN, FAHA
St. Vincent Mercy Hospital
Toledo, OH

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2015

Ad.3 **Month and Year of most recent revision:** 08, 2015

Ad.4 **What is your frequency for review/update of this measure?** biannual

Ad.5 **When is the next scheduled review/update for this measure?** 02, 2016

reprinted may not be up-to-date when used unless the copier or printer has verified the version to be up-to-date and affirms that, and 2) users participating in Joint Commission accreditation, including ORYX® vendors, are required to update their software and associated documentation based on the published manual production timelines.

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:** Continued from Section S.18 Calculation Algorithm/Measure Logic:

22. Check Arrival Date
a. If Arrival Date is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Arrival Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Arrival Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to check Arrival Time.

23. Check Arrival Time
a. If Arrival Time is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Arrival Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Arrival Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to the Timing III calculation.

24. Calculate Timing III. Timing III, in minutes, is equal to the Discharge Date and Discharge Time minus the Arrival Date and Arrival Time.
a. If the time in minutes is less than zero, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If the time in minutes is greater than or equal to zero and less than 360, the case will proceed to a Measure Category Assignment of B for overall rate CSTK-03 and will not be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If the time in minutes is greater than or equal to 360, continue processing and proceed to check Initial Hunt and Hess Scale Performed.

25. Check Initial Hunt and Hess Scale Performed
a. If Initial Hunt and Hess Scale Performed is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Initial Hunt and Hess Scale Performed equals No, continue processing and proceed to Step 29 and check Initial ICH Score Performed.
c. If Initial Hunt and Hess Scale Performed equals Yes, continue processing and proceed to check Initial Hunt and Hess Scale Date.

26. Check Initial Hunt and Hess Scale Date
a. If Initial Hunt and Hess Scale Date is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Initial Hunt and Hess Scale Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Initial Hunt and Hess Scale Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to check Initial Hunt and Hess Scale Time.

27. Check Initial Hunt and Hess Scale Time
a. If Initial Hunt and Hess Scale Time is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure

CSTK-03a and CSTK-03b.
b. If Initial Hunt and Hess Scale Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Initial Hunt and Hess Scale Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to the Timing IV calculation.

28. Calculate Timing IV. Timing IV, in minutes, is equal to the Initial Hunt and Hess Scale Date and Initial Hunt and Hess Scale Time minus the Arrival Date and Arrival Time.
a. If the time in minutes is less than zero, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If the time in minutes is greater than 360, the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If the time in minutes is greater than or equal to zero and less than or equal to 360, the case will proceed to a Measure Category Assignment of E for overall rate CSTK-03 and will be in the numerator population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.

29. Check Initial ICH Score Performed
a. If Initial ICH Score Performed is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Initial ICH Score Performed equals No, the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Initial ICH Score Performed equals Yes, continue processing and proceed to check Initial ICH Score Date.

30. Check Initial ICH Score Date
a. If Initial ICH Score Date is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Initial ICH Score Date equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Initial ICH Score Date equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to check Initial ICH Score Time.

31. Check Initial ICH Score Time
a. If Initial ICH Score Time is missing, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If Initial ICH Score Time equals Unable to Determine (UTD), the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If Initial ICH Score Time equals a Non-Unable to be Determined (non-UTD), continue processing and proceed to the Timing V calculation.

32. Calculate Timing V. Timing V, in minutes, is equal to the Initial ICH Score Date and Initial ICH Score Time minus the Arrival Date and Arrival Time.
a. If the time in minutes is less than zero, the case will proceed to a Measure Category Assignment of X for Overall Rate CSTK-03 and will be rejected. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
b. If the time in minutes is greater than 360, the case will proceed to a Measure Category Assignment of D for Overall Rate CSTK-03 and will be in the measure population. Continue processing and proceed to Step 33 to Initialize Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.
c. If the time in minutes is greater than or equal to zero and less than or equal to 360, the case will proceed to a Measure Category

Assignment of E for overall rate CSTK-03 and will be in the numerator population. Initialize the Measure Category Assignment for strata measure CSTK-03a and CSTK-03b.

33. Initialize the Measure Category Assignment for CSTK-03a and CSTK-03b equal to B. Do not change the Measure Category Assignment that was already calculated for the overall measure CSTK-03. The rest of the algorithm will reset the appropriate Measure Category Assignment to CSTK-03a and CSTK-03b.

34. Overall Rate Category Assignment
a. If Overall Rate CSTK-03 Category Assignment equals X, the case will proceed to a Measure Category Assignment of X and will not be in the Measure Population for strata measure CSTK-03a and CSTK-03b. Stop Processing.
b. If Overall Rate CSTK-03 Category Assignment equals B, the case will proceed to a Measure Category Assignment of B and will not be in the Measure Population for strata measure CSTK-03a and CSTK-03b. Stop Processing.
c. If Overall Rate CSTK-03 Category Assignment equals D or E, continue processing and proceed to check ICD-10-CM Principal or Other Diagnosis Codes.

35. Check ICD-10-CM Principal or Other Diagnosis Codes
a. If any ICD-10-CM Principal or Other Diagnosis Codes is on table 8.2a, set strata measure CSTK-03a equal to overall rate CSTK-03 Measure Category Assignment. Stop Processing.
b. If none ICD-10-CM Principal or Other Diagnosis Codes is on table 8.2a, continue processing and proceed to recheck ICD-10-CM Principal or Other Diagnosis Codes.

36. Recheck ICD-10-CM Principal or Other Diagnosis Codes
a. If any ICD-10-CM Principal or Other Diagnosis Codes is on table 8.2b, set strata measure CSTK-03b equal to overall rate CSTK-03 Measure Category Assignment. Stop Processing.

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2872
Measure Title: Dementia- Cognitive Assessment
Measure Steward: PCPI
Brief Description of Measure: Percentage of patients, regardless of age, with a diagnosis of dementia for whom an assessment of cognition is performed and the results reviewed at least once within a 12 month period.
Developer Rationale: Dementia is often characterized by the gradual onset and continuing cognitive decline in one or more domains including memory, executive function, language, judgment, and spatial abilities. (APA 2007) Cognitive deterioration represents a major source of morbidity and mortality and poses a significant burden on affected individuals and their caregivers.(NIH 2010)

More specifically, a number of studies have documented the significant impact that cognitive deterioration can have on patient outcomes, including greater risk for elder self-neglect, higher levels of disability and functional impairment, and earlier institutionalization. Elder self-neglect is a common and important public health issue across all socio-demographic and socioeconomic strata in the US with approximately 1.2 million cases of elder self-neglect reported annually in the US . A recent study suggests that elder self-neglect reported to social services agency was associated with increased risk of mortality, and there is a gradient relation between greater self-neglect severity and higher risk for mortality. Moreover, evidence indicates that reports of elder self-neglect to social services agencies are rising. The National Centers on Elder Abuse defines elder self-neglect as "…as the behavior of an elderly person that threatens his/her own health and safety. Self-neglect generally manifests itself in an older person as a refusal or failure to provide himself/herself with adequate food, water, clothing, shelter, personal hygiene, medication (when indicated), and safety precautions." In one well accepted model, impairment in cognitive function has been identified as representing one of the central factors associated with worsening vulnerability in the syndrome of elder self-neglect. In addition, decline in cognitive function combined with physical disability, lack of social network and inadequate support services magnify the inadequate ability for self-protection, leading to the syndrome of elder self-neglect. (Dong, Simon, Wilson et al 2010). Dyer and colleagues' have postulated that elder self-neglect occurs because of cognitive impairment secondary to a wide variety of medical etiologies, which subsequently impair the ability to perform activities of daily living. In particular, deficits in executive functioning are likely to promote self-neglect due to lack of insight and poor judgment in managing one's personal care needs (Wilkins, Horning, Castle et al 2014).

Krishna, Beulah, and Jones et al (2015) found that lower scores on individual domains of the 10/66 battery of cognitive tests were associated with higher levels of disability and functional impairment in community-dwelling older adults.

The most consistent predictors of nursing home admission in persons with dementia included severity of cognitive impairment, Alzheimer disease diagnosis, basic activity of daily living dependencies, behavioral symptoms, and depression. (Gaugler, Krichbaum, & Wyman 2009)

Patients who believe the end of life is near and who have a realistic understanding of the clinical problems characterizing terminal disease are more likely to receive care directed toward comfort. These observations extend to health care proxies for nursing home residents with advanced dementia (Mitchell, Teno, & Kiely et al 2009). Thus, measuring the

decline in cognition is important for knowing when dementia has advanced to its most severe stage.

Although cognitive deterioration follows a different course depending on the type of dementia, significant rates of decline have been reported. For example, one study found that the annual rate of decline for Alzheimer´s disease patients was more than four times that of older adults with no cognitive impairment (Wilson, Aggarwal, & Barnes 2010) Nevertheless, measurable cognitive abilities remain throughout the course of dementia.(APA 2007) Initial and ongoing assessments of cognition are fundamental to the proper management of patients with dementia. These assessments serve as the basis for identifying treatment goals, developing a treatment plan, monitoring the effects of treatment, and modifying treatment as appropriate and ensuring that care is aligned with the goals of management which are often focused on improving the quality of life for patients and caregivers, maintaining optimal function, reducing the rate of progression of functional and behavioral disturbances, and providing maximum comfort (Herrmann & Gauthier 2008).

APA 2007- American Psychiatric Association (APA). Practice guideline for the treatment of patients with Alzheimer´s disease and other dementias. Arlington (VA): American Psychiatric Association (APA); 2007 Oct.

Dong XQ; Simon MA; Wilson RS; Mendes de Leon CF; Rajan KB; Evans DA. Decline in Cognitive Function and Risk of Elder Self-Neglect: Finding from the Chicago Health Aging Project. J Am Geriatr Soc. 2010 Dec; 58(12): 2292–2299. doi: 10.1111/j.1532-5415.2010.03156.x
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3059228/

Gaugler JE, Yu F, Krichbaum K, Wyman JF. Predictors of nursing home admission for persons with dementia. Med Care. 2009 Feb;47(2):191-8. doi: 10.1097/MLR.0b013e31818457ce.
http://www.ncbi.nlm.nih.gov/pubmed/19169120

Herrmann N, Gauthier S. Diagnosis and treatment of dementia: 6. Management of severe Alzheimer disease. CMAJ. December 2, 2008; 179(12): 1279 – 1287

Krishna M, Beulah E, Jones S, Sundarachari R, A S, Kumaran K, Karat SC, Copeland JR, Prince M, Fall C. Cognitive function and disability in late life: an ecological validation of the 10/66 battery of cognitive tests among community-dwelling older adults in South India. Int J Geriatr Psychiatry. 2015 Dec 17. doi: 10.1002/gps.4404. [Epub ahead of print]
http://www.ncbi.nlm.nih.gov/pubmed/?term=krishna+10%2F66

Mitchell SL, Teno JM, Kiely DK, Shaffer ML, Jones RN, Prigerson HG, Volicer L, Givens JL, Hamel MB. The clinical course of advanced dementia. N Engl J Med. 2009 Oct 15;361(16):1529-38. doi: 10.1056/NEJMoa0902234.
http://www.nejm.org/doi/full/10.1056/nejmoa0902234#t=articleBackground

National Institutes of Health (NIH). NIH State-of-the-Science Conference: Preventing Alzheimer's Disease and Cognitive Decline. April 26–28, 2010. http://consensus.nih.gov/2010/docs/alz/alz_stmt.pdf. Accessed Jan 13, 2016.

Wilkins SS; Horning S; Castle S; Leff A; Hahn TJ; Chodosh J. Self-Neglect in Older Adults With Cognitive Impairment. Annals of Long Term Care- Clinical Care and Aging; Volume 22 - Issue 12 - December 2014
http://www.annalsoflongtermcare.com/article/challenges-and-management-self-neglect-older-adults-cognitive-impairment

Wilson RS, Aggarwal NT, Barnes LL, Mendes de Leon CF, Hebert LE, Evans DA. Cognitive decline in incident Alzheimer disease in a community population. Neurology. 2010 Mar 23;74(12):951-5.

Numerator Statement: Patients for whom an assessment of cognition is performed and the results reviewed at least once within a 12 month period
Definition:
Cognition can be assessed by the clinician during the patient´s clinical history. Cognition can also be assessed by direct examination of the patient using one of a number of instruments, including several originally developed and validated for

screening purposes. This can also include, where appropriate, administration to a knowledgeable informant. Examples include, but are not limited to:
-Blessed Orientation-Memory-Concentration Test (BOMC)
-Montreal Cognitive Assessment (MoCA)
-St. Louis University Mental Status Examination (SLUMS)
-Mini-Mental State Examination (MMSE) [Note: The MMSE has not been well validated for non-Alzheimer´s dementias]
-Short Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE)
-Ascertain Dementia 8 (AD8) Questionnaire
-Minimum Data Set (MDS) Brief Interview of Mental Status (BIMS) [Note: Validated for use with nursing home patients only]
-Formal neuropsychological evaluation
-Mini-Cog
Denominator Statement: All patients, regardless of age, with a diagnosis of dementia
Denominator Exclusions: Exceptions: Documentation of medical reason(s) for not assessing cognition (eg, patient with very advanced stage dementia, other medical reason)
Documentation of patient reason(s) for not assessing cognition

Measure Type: Process
Data Source: Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record
Level of Analysis: Clinician : Group/Practice, Clinician : Individual, Clinician : Team

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

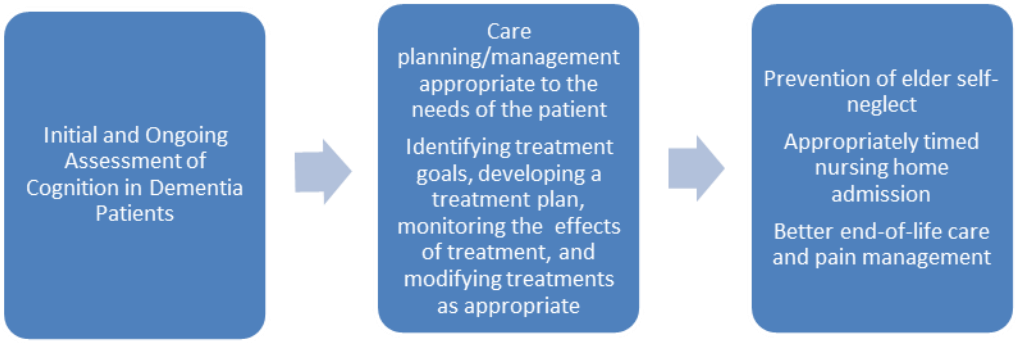
1a. Evidence

1a. Evidence. The evidence requirements for a *process or intermediate outcome* measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured.

The developer provides the following evidence for this measure:

- **Systematic Review of the evidence specific to this measure?** ☒ **Yes** ☐ **No**
- **Quality, Quantity and Consistency of evidence provided?** ☒ **Yes** ☐ **No**
- **Evidence graded?** ☒ **Yes** ☐ **No**

The developer provides the following diagram to support the relationship between the process of care (assessment for cognition) and outcomes:



Two clinical guideline sources are cited to support the measure; the American Psychiatric Association (APA) guidelines state the following: *Ongoing assessment includes periodic monitoring of the development and evolution of cognitive and noncognitive psychiatric symptoms and their response to intervention [I].* Additionally, the California Workgroup on Guidelines for Alzheimer's Disease Management recommendation is to: *Conduct and document an assessment and monitor changes in cognitive status using a reliable and valid instrument.* The California recommendation is not rated, however, the APA [I] rating indicates: Recommended with substantial clinical confidence.

The APA Guidelines cite their basis of evidence as a cohort or longitudinal study. A study in which subjects are prospectively followed over time without any specific intervention. The California guidelines do not indicate evidence base. The developer summarized 7 additional studies regarding the impact of diminished cognitive function in dementia/Alzheimer's patients. These studies were not specific to the relationship between screening and better outcomes, but do substantiate that declining cognition impacts activities of daily living, and knowledge of level of cognition and functional capacity can assist in targeted interventions.

Exception to evidence
N/A

Guidance from the Evidence Algorithm
Process measure/systematic review (Box 3) →QQC not presented (Box 4) →Quantity: Moderate; Quality: Moderate (cohort studies and "B" clinical guideline recommendation); Consistency: Moderate →Box 6 Moderate

Questions for the Committee:

- *This is a process measure and the developer should establish a relationship between the intervention and the patient outcome. The literature is sparse in this area, but there is a strong clinical guideline. Do you agree the evidence indicates with high certainty that the benefits of assessing for cognition outweigh undesirably effects?*
  - *How strong is the evidence for this relationship?*
  - *Is the evidence directly applicable to the process of care being measured?*

Preliminary rating for evidence: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities
Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The performance scores were provided by a national network of 30 post acute care facilities, across the United States. The sites within this network provide skilled nursing, long term care, rehabilitation and memory care to post acute care patients.

- 2015 data with 15,691 encounters with Dementia patients, showed a mean performance rate of 63.93%.
- In addition, a 2007 analysis of medical records and caregiver surveys for 378 patients with dementia found that only 50% of patients received an assessment of their cognitive status in the previous 12 months.(1) Another study surveying clinicians practicing in VA medical centers found that only two thirds of clinicians reported regularly performing a standardized assessment of cognitive functioning.(2) Despite the availability and dissemination of established best practice guidelines, there is still wide variation in physician practice patterns in dementia care. The quality of currently available studies limits the ability to draw strong conclusions. (3)

Disparities

The developer stated the following: we are not aware of any publications/evidence outlining disparities specifically related to assessing cognition in dementia patients. However, a systematic review and meta-analysis of the use of dementia treatment, care, and research identified significant racial and ethnic disparities in western countries, particularly the United States. Overall, the authors found "consistent evidence, mostly from the United States, that [minority ethnic] people accessed diagnostic services later in their illness, and once they received a diagnosis, were less likely to access anti-dementia medication, research trials, and 24-hour care."

Questions for the Committee:

- *Do you think the data provided reflects the current standard of care?*
- *Do you agree the data reported indicates a gap in care that warrants a national performance measure?*
- *If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

*This is a process measure. The developers have provided evidence to support the facts that dementia is associated with increased morbidity & mortality and that individuals with Alzheimers disease experience a more rapid decline of cognition compared to others. They argue that information about an individual's cognition can be used to help provide optimal care. There is not evidence, however, that EXAMINING cognition will actually impact this care...although that would seem to be consistent with expert opinion. The developers cite 2 CPGs that recommend this practice.

*This process measure offers potential to identify cognitive declines that will result in declines in self-care with expected increased in poorer health outcomes. The data are not terribly extensive, but generally persuasive that unmonitored cognitive declines lead to poorer outcomes.

*Evidence supports benefits of assessing cognition in dementia pts

1b. Performance Gap

*"The developers provide examples to demonstrate what would seem to be sub-optimal performance in this area (50-64%). Since I believe the importance of the measure has been justified, measuring % adherence with an overall target to improve the percentage seems warranted. Disparity data presented is not specific enough to be a major factor (e.g. disparities in dementia care in western countries, esp the US and the fact that minority/ethnic individuals access care later in disease and are less likely to receive optimal care)"

*Data are presented indicating modest regular collection of data on cognitive status. Some additional data suggest that this may be less common in some populations. The data linking the absence of regular cognitive monitoring to poorer care is more inferential than concrete. Since this measure is seeking approval for trial use this seem appropriate.

*Data support gap in care that warrants the measure

1c. High Priority (previously referred to as High Impact)

*The condition prevalence and impact have been well established. Congruency with public health priorities is discussed. Opportunities for performance improvement are identified with the significance explained.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

Maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- This is a new eMeasure – HQMF specification are included. See eMeasure Technical Review below.

Questions for the Committee :

- *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor review

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF). **HQMF specifications** ☒ **Yes** ☐ **No** **eMeasure Trial Approval Technical Review has found that:** • The submitted eMeasure specification captures the data elements and measure logic needed for automated measure calculation
Documentation of HQMF or QDM limitations	All components in the measure logic of the submitted eMeasure are represented using the HQMF and QDM.
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously
Feasibility Testing	The feasibility analysis submitted by the measure developer meets the requirements to be considered for eMeasure Trial Approval.

2a2. Reliability Testing Testing attachment
Maintenance measures – less emphasis if no new testing data provided

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

Initial reliability testing was conducted in the Bonnie test deck; the overall patient simulation included 32 patients, with a diagnosis of Dementia & Mental Degenerations. 3 patients had a diagnosis of Severe Dementia. As a measure under consideration for the Trial Approval program, the developers must indicate if they have a plan in place for full testing (reliability and validity) and this information will be submitted and evaluated by NQF prior to any consideration of full measure endorsement. The Testing attachment indicates a plan for reliability and validity testing.

Questions for the Committee:

- *The Committee will not be asked to vote on Reliability for this emeasure since it is being considered for Trial Use; however, questions regarding the testing plan and other concerns about reliability are welcome for discussion.*

Preliminary rating for reliability: ☐ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Full specifications are not provided; however, numerator, denominator and exclusions are provided for consideration of the measure construct.

Question for the Committee:

- *Based on the information provided, and intent of the measure to assess for cognitive function in the dementia population – do you feel the specifications are consistent with evidence? Are there any exclusions you would recommend for consideration?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

The only testing completed to date includes Bonnie testing and some review for feasibility. This measure is being considered for trial use, thus full validity testing results are not expected and the Committee will not vote on this criterion.

Questions for the Committee:

- *Other specific question of the validity testing?*

2b3-2b7. Threats to Validity

2b3. Exclusions:
The following information was provided: For measure Dementia: Cognitive Assessment, exceptions may include medical reason(s) (eg, patients with very advanced stage dementia, other medical reason) or patient reason(s) for not assessing cognition. Where examples of exceptions are included in the measure language, value sets for these examples are developed and included in the eMeasure. Although this methodology does not require the external reporting of more detailed exception data, the PCPI recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The PCPI also advocates the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

Questions for the Committee:

- *Are there other threats to validity the measure developer should consider?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☐ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary This will not be evaluated by the Committee; the developer has indicated no risk adjustment will be used for this measure.

Questions for the Committee:

- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

Unknowns at this time.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

Extent not known at this time.

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

*The only question I have here is relative to exclusions. I'm wondering if the "out" to doing the assessment is too unspecified and if the barriers to capturing this in the text search of the EHR is reasonable.
*We are told that reliability will not be rated for this measure. Since the instruments used for the cognitive assessments are all independently evaluated for reliability this should be no concern at this time.
*No concerns
*This measure is being considered for trial use. I believe this means that this element will not be judged at this time.
*We are told that validity will not be rated for this measure. I do have a concern that is not addressed in the proposal. Since the presumed goal is the ability to track change in cognitive (presumably declines) the potential that separate measures could be used which would not allow any direct comparison could limit the utility of the measure.
*Definition of assessment is broad, eg, clinical hx or screening/direct testing

2a2. Reliability Testing

*Reliability testing was conducted with a patient simulation including 32 patients. As I am not familiar with the Bonnie test deck, I am not confident to comment on the adequacy of this test.

2b2. Validity Testing

*as above
*See above

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

*Exclusions from the denominator of those groups identified seems entirely appropriate and no threat to the utility of the measure. Presumably assessment over time with this population would have an end point where a person's condition is so severe as to warrant exclusion from further testing. The precise criteria (as opposed to the general grounds) for exclusion are not identified. Perhaps at some point a score below X in a cognitive test would become grounds for exclusion rather than just basic it on clinical judgment.
*Many unknowns at this stage of pilot testing

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

Feasibility analysis submitted by the measure developer meets the requirements to be considered for eMeasure Trial Approval. While the findings of the eMeasure Technical Review indicate that this measure has not been adequately tested for NQF endorsement and use in quality improvement and accountability applications, this measure is ready for implementation in real world settings for the purpose of additional testing. Based on this review, the submitted eMeasure specification is capable of being processed and interpreted by clinical information systems.

The developer indicates: A 2016 feasibility assessment was performed in order to assess the extent to which the required data are readily available or could be captured without undue burden and can be implemented for performance measurement.

Two entities participated in the Feasibility assessment for this measure.
- One participating entity is an Electronic Health Record (EHR) vendor that incorporates this measure as part of the Dementia Measures Group, included in the Physician Quality Reporting System program. The vendor's domain is primarily work with nursing home facilities.
- The second participating entity is a national network of 30 post‐acute care facilities, across the United States. The sites within this network provide skilled nursing, long term care, rehabilitation and memory care to post‐acute care patients.
- Two different EHR products were included in this assessment.

Feasibility Assessment results
All data elements for this measure are captured in the Electronic Health Record. The measure exceptions are the only data element not currently being captured in a structured format within the two EHR products assessed for feasibility. Exception information is extracted from documentation within the medical record, explaining why the patient did not receive the standard of care. Although this information is currently entered as text, it is identifiable in the medical record. Future iterations of vendor products will likely allow health care providers to collect exceptions information in a structured format.

Questions for the Committee:
- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *If an eMeasure, does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes
3b. Electronic Sources
3c. Data Collection Strategy
*"Feasibility was tested in 2 EHRs. As above, there is a concern about the text extraction of date about exclusions."
*Given the normal inclusion of the appropriate information in the EHR, and this as a proposed eMeasure, this seems quite feasible.
*Testing was in networks of nursing home/post-acute care facilities probably with MDS data elements to assess dementia. How will this translate to the acute care setting with different EHR data elements?

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure [from OPUS]
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No**
OR
Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details This measure is in current use in PQRS and Meaningful Use Stage 2

Improvement results The developer indicates that although the measure is in public use, performance results are not available.

Unexpected findings (positive or negative) during implementation None indicated

Potential harms None indicated

Feedback :

None indicated

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

*"Measure is in use in PQRS & Meaningful Use Stage 2. According to developer, performance results are not available. Monitoring adherence to measure is in line with CPG best practice recommendations and is likely to impact care...or at least provide information that can be further analyzed to more directly impact (and judge the impact of) care. "

*Data are easily reported and may have considerable potential to to both refine treatment in patients with dementia and allow more precise monitoring of the impact of potential ameliorative treatments.

*Performance results unavailable

Criterion 5: Related and Competing Measures

Related or competing measures

None

Harmonization

N/A

Pre-meeting public and member comments

Comment by Amy Elaine Sanders, MD
Organization American Academy of Neurology
Comment #5573: favor endorsement

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Dementia- Cognitive Assessment
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 1/15/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome

☐ Health outcome: Click here to name the health outcome

☐ Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

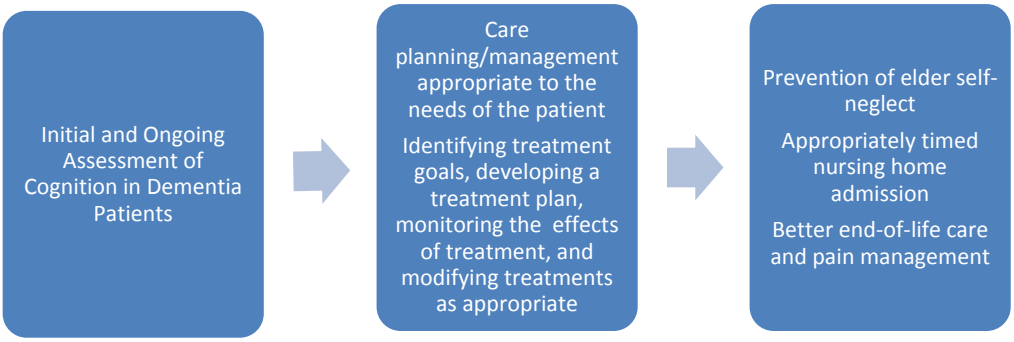
☒ Process: Measuring Cognition using a validated tool

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☒ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***

☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***

☒ Other – ***complete section 1a.8***

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

American Psychiatric Association (APA). Practice guideline for the treatment of patients with Alzheimer´s disease and other dementias. Arlington (VA): American Psychiatric Association (APA); 2007 Oct.
http://psychiatry.org/File%20Library/Psychiatrists/Practice/Clinical%20Practice%20Guidelines/alzheimers.pdf

APA Guideline Watch (2014 update of literature)- published only on web-site
http://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/alzheimerwatch.pdf

Published only on web-site
https://www.cdph.ca.gov/programs/alzheimers/Documents/professional_GuidelineFullReport.pdf

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

APA

P11- GENERAL TREATMENT PRINCIPLES AND ALTERNATIVES/ Psychiatric Management

Ongoing assessment includes periodic monitoring of the development and evolution of cognitive and noncognitive psychiatric symptoms and their response to intervention [I]

Supporting text:

P17-18, II. FORMULATION AND IMPLEMENTATION OF ATREATMENT PLAN; B. PSYCHIATRIC MANAGEMENT; 3. Assess and Monitor Psychiatric Status

Cognitive symptoms that almost always require assessment include impairments in memory, executive function, language, judgment, and spatial abilities. It is often helpful to track cognitive status with a structured simple examination. If the same instrument is used repeatedly, the clinician should watch for practice effects. A detailed assessment of functional status may also aid the clinician in documenting and tracking changes over time as well as providing guidance to the patient and caregivers. Functional status is typically described in terms of the patient's ability to perform instrumental activities of daily living such as shopping, writing checks, basic housework, and activities of daily living such as dressing, bathing, feeding, transferring, and maintaining continence. These regular assessments of recent cognitive and functional status provide a baseline for assessing the effect of any intervention, and they improve the recognition and treatment of acute problems, such as delirium.

California Workgroup on Guidelines for Alzheimer's Disease Management

P9- Assessment- Assessment: Cognitive Status

Recommendation: Conduct and document an assessment and monitor changes in cognitive status using a reliable and valid instrument.

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

APA- [I] Recommended with substantial clinical confidence

CA Working Group- no grade assigned

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

APA-

Each recommendation is identified as falling into one of three categories of endorsement, indicated by a bracketed Roman numeral following the statement. The three categories represent varying levels of clinical confidence:

[I] Recommended with substantial clinical confidence

[II] Recommended with moderate clinical confidence

[III] May be recommended on the basis of individual circumstances

CA Working Group-

No rating of recommendations is included

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☒ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

Complete section 1a.7

_____

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

Complete section 1a.7

_____

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

The initial evidence review from the APA guideline searched for: PubMed- "dementia," "dementias," "Alzheimer," "Alzheimer's," "Pick disease," or "mild cognitive impairment

The Guideline Watch update searched:
PubMed MeSH terms "Alzheimer Disease," "Creutzfeldt-Jakob Syndrome," "Dementia," "Dementia, Multi-Infarct," "Dementia, Vascular," "Lewy Body Disease," "Pick Disease of the Brain," and "Cognition Disorders" as well as the following title and abstract words or phrases: "Alzheimer," "Alzheimer's," "CADASIL," "cortical dementia," "cortical dementias," "dementia with lewy bodies," "dementia," "dementias," "frontotemporal dementia," "lewy body dementia," "mild cognitive impairment," "Parkinson's dementia," "subcortical dementia," "subcortical dementias," "vascular dementia," "vascular dementias," "cognitive disorder," "cognitive disorders," "cognitive impairment," or "cognitive impairments." Titles, abstracts, and keywords in the Cochrane database were searched for the words "dementia," "dementias," and "cognitive."

There is no documentation of the evidence review for the CA Working Group guideline.

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

APA C- *Cohort or longitudinal study.* A study in which subjects are prospectively followed over time without any specific intervention.

CA Working Group- no grade assigned

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

APA-

The following coding system is used to indicate the nature of the supporting evidence in the references:
[A] *Double-blind, randomized clinical trial.* A study of an intervention in which subjects are prospectively followed over time; there are treatment and control groups; subjects are randomly assigned to the two groups; both the subjects and the investigators are blind to the assignments.
[A–] *Randomized clinical trial.* Same as above, but not double-blind.
[B] *Clinical trial.* A prospective study in which an intervention is made and the results of that intervention are tracked longitudinally; study does not meet standards for a randomized clinical trial.
[C] *Cohort or longitudinal study.* A study in which subjects are prospectively followed over time without any specific intervention.
[D] *Case-control study.* A study in which a group of patients is identified in the present and information about them is pursued retrospectively or backward in time.
[E] *Review with secondary data analysis.* A structured analytic review of existing data, for example, a meta-analysis or a decision analysis.
[F] *Review.* A qualitative review and discussion of previously published literature without a quantitative synthesis of the data.
[G] *Other.* Textbooks, expert opinions, case reports, and other reports not included above.

CA Working Group- no grading system documented

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: 1994-2012

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

Information regarding the total number of studies and type of study designs included in the body of evidence is not available.

However, the APA guideline cites 2 Level C studies in support of the recommendation statement.

The CA Working Group cites only studies supporting the validation of the tools used for cognitive assessment.

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

Information regarding the overall quality of evidence across studies is not available.

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

The guideline does not include an overall estimate of benefit from the body of evidence. However, they do include the following summary information regarding the benefits of initial and ongoing cognitive assessments in patients with dementia, “These regular assessments of recent cognitive and functional status provide a baseline for assessing the effect of any intervention, and they improve the recognition and treatment of acute problems, such as delirium.”

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

The guideline did not refer to any harms associated with a cognitive assessment.

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for each new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

See 1a.8

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

The APA guideline was published in 2007 and CA Working Group guideline in 2008, prior to the release of the IOM standards for clinical practice guidelines or systematic reviews. The ratings of the recommendations in the APA guideline were based on what APA deemed "clinical confidence" and studies to support the concept of cognitive assessment were limited in the review. Thus, we undertook a literature search of our own to identify further studies that support this measure.

We searched PubMed and Google Scholar, from 2008-2016 using the terms:
Dementia AND "cognitive assessment," "cognitive impairment," "cognitive deterioration," "decline in cognition." "Cognitive decline and self-neglect." "cognitive decline AND alzheimer's disease."

Titles and abstracts were reviewed to ascertain whether studies reviewed the importance of measuring cognition in dementia patients. Outcomes of declining cognition are described in the rationale.

1a.8.2. Provide the citation and summary for each piece of evidence.

1) **National Institutes of Health (NIH). NIH State-of-the-Science Conference: Preventing Alzheimer's Disease and Cognitive Decline. April 26–28, 2010. http://consensus.nih.gov/2010/docs/alz/alz_stmt.pdf. Accessed Jan 13, 2016.**

Cognitive decline and Alzheimer's disease are major causes of morbidity and mortality worldwide and are substantially burdensome to the affected persons, their caregivers, and society in general. Extensive research over the past 20 years has provided important insights on the nature of Alzheimer's disease and cognitive decline and the magnitude of the problem. Nevertheless, there remain important and formidable challenges in conducting research on these diseases, particularly in the area of prevention. Currently, firm conclusions cannot be drawn about the association of any modifiable risk factor with cognitive decline or Alzheimer's disease. Highly reliable consensus-based diagnostic criteria for cognitive decline, mild cognitive impairment, and Alzheimer's disease are lacking, and available criteria have not been uniformly applied. Evidence is insufficient to support the use of pharmaceutical agents or dietary supplements to prevent cognitive decline or Alzheimer's disease.

2) **Dong XQ; Simon MA; Wilson RS; Mendes de Leon CF; Rajan KB; Evans DA. Decline in Cognitive Function and Risk of Elder Self-Neglect: Finding from the Chicago Health Aging Project. J Am Geriatr Soc. 2010 Dec; 58(12): 2292–2299. doi: 10.1111/j.1532-5415.2010.03156.x**

Objectives
This study aimed to examine the longitudinal association between decline in cognitive function and risk of elder self-neglect in a community-dwelling population.

Design
Prospective population-based study

Setting
Geographically-defined community in Chicago.

Participants
Community-dwelling subjects reported to the social services agency from 1993–2005 for self-neglect who also participated in the Chicago Health Aging Project (CHAP). Of the 5,519 participants in the Chicago Health Aging Project, 1,017 were reported to social services agency for suspected elder self-neglect from 1993–2005.

Measurements
Reported elder self-neglect was identified by social services agency. The primary predictor was decline in cognitive function assessed using the Mini-Mental State Examination (MMSE), the Symbol Digit Modalities Test (Executive Function), and both immediate and delayed recall of the East Boston Memory Test (Episodic Memory). An index of global

cognitive function scores was derived by averaging z-scores of all tests. Outcome of interest was elder self-neglect. Logistic and linear regression models were used to assess these longitudinal associations.

Results
After adjusting for potential confounding factors, decline in global cognitive function, MMSE or episodic memory was not independently associated with increased risk of reported and confirmed elder self-neglect. Decline in executive function was associated with increased risk of reported and confirmed elder self-neglect. Decline in global cognitive function was associated with increased risk of greater self-neglect severity (PE=0.76, SE=0.31, p=0.014).
Conclusion
Decline in executive function was associated with increased risk of reported and confirmed elder self-neglect. Decline in global cognitive function was associated with increased risk of greater self-neglect severity.

3) **Wilkins SS; Horning S; Castle S; Leff A; Hahn TJ; Chodosh J. Self-Neglect in Older Adults With Cognitive Impairment. Annals of Long Term Care- Clinical Care and Aging; Volume 22 - Issue 12 - December 2014**

Older adults with major neurocognitive impairment (eg, dementia) and poor insight due to executive/frontal deficits often do not understand their lack of ability to safely perform instrumental activities of daily living (IADLs; eg, medication management, finances, driving). This can lead to self-neglect, a problem that is commonly encountered in geriatric practice in patients with cognitive impairment. An interdisciplinary approach is useful in managing these complicated cases. Cognitive and capacity assessment regarding IADL concerns and options for self-care are required, as well as potential involvement of Adult Protective Services and probate conservatorship. The authors discuss common issues encountered in managing these cases and illustrate the process through two case presentations.

4) **Krishna M, Beulah E, Jones S, Sundarachari R, A S, Kumaran K, Karat SC, Copeland JR, Prince M, Fall C. Cognitive function and disability in late life: an ecological validation of the 10/66 battery of cognitive tests among community-dwelling older adults in South India. Int J Geriatr Psychiatry. 2015 Dec 17. doi: 10.1002/gps.4404. [Epub ahead of print]**

BACKGROUND:
The 10/66 Dementia Research Group developed and validated a culture and education fair battery of cognitive tests for diagnosis of dementia in population-based studies in low-income and middle-income countries including India.

AIMS:
This study examined the association between individual domains of the 10/66 battery of cognitive tests and 'disability' and 'functional impairment' in community-dwelling older adults in South India.

METHODS:
One hundred twenty-nine adults aged 60-90 years residing in Karunapura, in the city of Mysore, were interviewed in their own homes. Cognitive functioning was measured by administering the 10/66 battery of cognitive tests that composes of Community Screening Instrument for Dementia (CSI'D' COGSCORE), verbal fluency (VF) and word list memory recall (WLMR). A reliable informant was interviewed to ascertain if the subject's cognitive problems have resulted in functional impairment. Disability was measured by WHO Disability Schedule-II (DAS).

RESULTS:
The women had significantly lower CSI'D' COGSCORE score when compared with men ($p = 0.002$). The presence of 'functional impairment' resulting from cognitive decline was significantly associated with lower scores on VF ($p = 0.03$), WLMR ($p = 0.03$) and CSI'D' COGSCOREs ($p < 0.01$). There was a significant inverse association between WHO DAS II score and WLMR ($p = 0.004$), VF (0.006) and CSI'D' COGSCORE scores ($p \leq 0.001$) even after adjusting for self-reported ischaemic heart disease, stroke, chronic obstructive airway disease, hypertension and diabetes.

CONCLUSIONS:
Lower scores on individual domains of the 10/66 battery of cognitive tests are associated with higher levels of disability and functional impairment in community-dwelling older adults. These culture and education fair tests are suitable for use in population-based research in India.

5) **Mitchell SL, Teno JM, Kiely DK, Shaffer ML, Jones RN, Prigerson HG, Volicer L, Givens JL, Hamel MB. The clinical course of advanced dementia. N Engl J Med. 2009 Oct 15;361(16):1529-38. doi: 10.1056/NEJMoa0902234.**

BACKGROUND:
Dementia is a leading cause of death in the United States but is underrecognized as a terminal illness. The clinical course of nursing home residents with advanced dementia has not been well described.
METHODS:

We followed 323 nursing home residents with advanced dementia and their health care proxies for 18 months in 22 nursing homes. Data were collected to characterize the residents' survival, clinical complications, symptoms, and treatments and to determine the proxies' understanding of the residents' prognosis and the clinical complications expected in patients with advanced dementia.

RESULTS:
Over a period of 18 months, 54.8% of the residents died. The probability of pneumonia was 41.1%; a febrile episode, 52.6%; and an eating problem, 85.8%. After adjustment for age, sex, and disease duration, the 6-month mortality rate for residents who had pneumonia was 46.7%; a febrile episode, 44.5%; and an eating problem, 38.6%. Distressing symptoms, including dyspnea (46.0%) and pain (39.1%), were common. In the last 3 months of life, 40.7% of residents underwent at least one burdensome intervention (hospitalization, emergency room visit, parenteral therapy, or tube feeding). Residents whose proxies had an understanding of the poor prognosis and clinical complications expected in advanced dementia were much less likely to have burdensome interventions in the last 3 months of life than were residents whose proxies did not have this understanding (adjusted odds ratio, 0.12; 95% confidence interval, 0.04 to 0.37).

CONCLUSIONS:
Pneumonia, febrile episodes, and eating problems are frequent complications in patients with advanced dementia, and these complications are associated with high 6-month mortality rates. Distressing symptoms and burdensome interventions are also common among such patients. Patients with health care proxies who have an understanding of the prognosis and clinical course are likely to receive less aggressive care near the end of life.

6) **Gaugler JE, Yu F, Krichbaum K, Wyman JF. Predictors of nursing home admission for persons with dementia. Med Care. 2009 Feb;47(2):191-8. doi: 10.1097/MLR.0b013e31818457ce.**

OBJECTIVE:
The objective of this systematic review was to identify factors that consistently predict nursing home admission (NHA) in persons with dementia.

METHODS:
Studies published in English were retrieved by searching the MEDLINE (1966-2006), PSYCINFO (1950-2006), CINAHL (1982-2006), and Digital Dissertations (1950-2006) databases. Bibliographies of retrieved studies were also searched. Information on study characteristics and empirical results were extracted using a standardized protocol.

RESULTS:
Of 782 relevant studies identified 80 were selected for review based upon eligibility criteria. The most consistent predictors of NHA in persons with dementia included severity of cognitive impairment, Alzheimer disease diagnosis, basic activity of daily living dependencies, behavioral symptoms, and depression. Caregivers who indicated greater emotional stress, a desire to institutionalize the care recipient, and feelings of being "trapped" in care responsibilities were more likely to admit persons with dementia to nursing homes. Demographic variables, incontinence, and service use did not consistently predict NHA.

CONCLUSIONS:
Several results seemed to challenge conventional assumptions of what precipitates NHA among persons with dementia. Caregiver stressors in conjunction with care recipient characteristics are important to consider when assessing NHA risk. The findings emphasize the need to construct more complex models of institutionalization when designing risk measures to target interventions.

7) **Wilson RS1, Aggarwal NT, Barnes LL, Mendes de Leon CF, Hebert LE, Evans DA. Cognitive decline in incident Alzheimer disease in a community population. Neurology. 2010 Mar 23;74(12):951-5. doi: 10.1212/WNL.0b013e3181d64786.**

OBJECTIVE:
To measure the cognitive consequences of incident Alzheimer disease (AD) in older African American and white subjects.

METHODS:
Data are from the Chicago Health and Aging Project, a longitudinal cohort study of older white and black persons residing in a geographically defined community. At 3-year intervals, the entire study population completed 4 brief cognitive tests, from which a previously established composite measure of global cognition was derived, and a subset underwent detailed clinical evaluation that supported clinical classification of mild cognitive impairment, dementia, and AD. We used mixed-effects models to examine change in cognitive function following the diagnostic evaluation.

RESULTS:

On clinical evaluation, 614 persons were found to have no cognitive impairment, 395 had mild cognitive impairment, and 149 had AD (88.5% mild); 10 persons with other dementias were excluded from analyses. During up to 11 years of observation following the clinical evaluation (mean = 5.5, SD = 2.5), the composite measure of global cognition declined a mean of 0.042 unit per year (SE = 0.008, $p < 0.001$) in those with no cognitive impairment. In comparison to the no cognitive impairment group, the annual rate of decline was increased more than twofold in mild cognitive impairment (estimate = 0.086, SE = 0.011, $p < 0.001$) and more than fourfold in AD (estimate = 0.173, SE = 0.020, $p < 0.001$). Results did not reliably vary by race, sex, or age.

CONCLUSIONS:

Alzheimer disease has a devastating impact on cognition, even in its prodromal stages, with comparable effects in African American and white persons.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
evidence_attachment_dementia_cognitive_assessment_rev.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Dementia is often characterized by the gradual onset and continuing cognitive decline in one or more domains including memory, executive function, language, judgment, and spatial abilities. (APA 2007) Cognitive deterioration represents a major source of morbidity and mortality and poses a significant burden on affected individuals and their caregivers.(NIH 2010)

More specifically, a number of studies have documented the significant impact that cognitive deterioration can have on patient outcomes, including greater risk for elder self-neglect, higher levels of disability and functional impairment, and earlier institutionalization. Elder self-neglect is a common and important public health issue across all socio-demographic and socioeconomic strata in the US with approximately 1.2 million cases of elder self-neglect reported annually in the US . A recent study suggests that elder self-neglect reported to social services agency was associated with increased risk of mortality, and there is a gradient relation between greater self-neglect severity and higher risk for mortality. Moreover, evidence indicates that reports of elder self-neglect to social services agencies are rising. The National Centers on Elder Abuse defines elder self-neglect as "…as the behavior of an elderly person that threatens his/her own health and safety. Self-neglect generally manifests itself in an older person as a refusal or failure to provide himself/herself with adequate food, water, clothing, shelter, personal hygiene, medication (when indicated), and safety precautions." In one well accepted model, impairment in cognitive function has been identified as representing one of the central factors associated with worsening vulnerability in the syndrome of elder self-neglect. In addition, decline in cognitive function combined with physical disability, lack of social network and inadequate support services magnify the inadequate ability for self-protection, leading to the syndrome of elder self-neglect. (Dong, Simon, Wilson et al 2010). Dyer and colleagues' have postulated that elder self-neglect occurs because of cognitive impairment secondary to a wide variety of medical etiologies, which subsequently impair the ability to perform activities of daily living. In particular, deficits in executive functioning are likely to promote self-neglect due to lack of insight and poor judgment in managing one's personal care needs (Wilkins, Horning, Castle et al 2014).

Krishna, Beulah, and Jones et al (2015) found that lower scores on individual domains of the 10/66 battery of cognitive tests were associated with higher levels of disability and functional impairment in community-dwelling older adults.

The most consistent predictors of nursing home admission in persons with dementia included severity of cognitive impairment, Alzheimer disease diagnosis, basic activity of daily living dependencies, behavioral symptoms, and depression. (Gaugler, Krichbaum, & Wyman 2009)

Patients who believe the end of life is near and who have a realistic understanding of the clinical problems characterizing terminal disease are more likely to receive care directed toward comfort. These observations extend to health care proxies for nursing home residents with advanced dementia (Mitchell, Teno, & Kiely et al 2009). Thus, measuring the decline in cognition is important for knowing when dementia has advanced to its most severe stage.

Although cognitive deterioration follows a different course depending on the type of dementia, significant rates of decline have been reported. For example, one study found that the annual rate of decline for Alzheimer´s disease patients was more than four times that of older adults with no cognitive impairment (Wilson, Aggarwal, & Barnes 2010) Nevertheless, measurable cognitive abilities remain throughout the course of dementia.(APA 2007) Initial and ongoing assessments of cognition are fundamental to the proper

management of patients with dementia. These assessments serve as the basis for identifying treatment goals, developing a treatment plan, monitoring the effects of treatment, and modifying treatment as appropriate and ensuring that care is aligned with the goals of management which are often focused on improving the quality of life for patients and caregivers, maintaining optimal function, reducing the rate of progression of functional and behavioral disturbances, and providing maximum comfort (Herrmann & Gauthier 2008).

APA 2007- American Psychiatric Association (APA). Practice guideline for the treatment of patients with Alzheimer´s disease and other dementias. Arlington (VA): American Psychiatric Association (APA); 2007 Oct.

Dong XQ; Simon MA; Wilson RS; Mendes de Leon CF; Rajan KB; Evans DA. Decline in Cognitive Function and Risk of Elder Self-Neglect: Finding from the Chicago Health Aging Project. J Am Geriatr Soc. 2010 Dec; 58(12): 2292–2299. doi: 10.1111/j.1532-5415.2010.03156.x
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3059228/

Gaugler JE, Yu F, Krichbaum K, Wyman JF. Predictors of nursing home admission for persons with dementia. Med Care. 2009 Feb;47(2):191-8. doi: 10.1097/MLR.0b013e31818457ce.
http://www.ncbi.nlm.nih.gov/pubmed/19169120

Herrmann N, Gauthier S. Diagnosis and treatment of dementia: 6. Management of severe Alzheimer disease. CMAJ. December 2, 2008; 179(12): 1279 – 1287

Krishna M, Beulah E, Jones S, Sundarachari R, A S, Kumaran K, Karat SC, Copeland JR, Prince M, Fall C. Cognitive function and disability in late life: an ecological validation of the 10/66 battery of cognitive tests among community-dwelling older adults in South India. Int J Geriatr Psychiatry. 2015 Dec 17. doi: 10.1002/gps.4404. [Epub ahead of print]
http://www.ncbi.nlm.nih.gov/pubmed/?term=krishna+10%2F66

Mitchell SL, Teno JM, Kiely DK, Shaffer ML, Jones RN, Prigerson HG, Volicer L, Givens JL, Hamel MB. The clinical course of advanced dementia. N Engl J Med. 2009 Oct 15;361(16):1529-38. doi: 10.1056/NEJMoa0902234.
http://www.nejm.org/doi/full/10.1056/nejmoa0902234#t=articleBackground

National Institutes of Health (NIH). NIH State-of-the-Science Conference: Preventing Alzheimer's Disease and Cognitive Decline. April 26–28, 2010. http://consensus.nih.gov/2010/docs/alz/alz_stmt.pdf. Accessed Jan 13, 2016.

Wilkins SS; Horning S; Castle S; Leff A; Hahn TJ; Chodosh J. Self-Neglect in Older Adults With Cognitive Impairment. Annals of Long Term Care- Clinical Care and Aging; Volume 22 - Issue 12 - December 2014
http://www.annalsoflongtermcare.com/article/challenges-and-management-self-neglect-older-adults-cognitive-impairment

Wilson RS, Aggarwal NT, Barnes LL, Mendes de Leon CF, Hebert LE, Evans DA. Cognitive decline in incident Alzheimer disease in a community population. Neurology. 2010 Mar 23;74(12):951-5.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
The performance scores were provided by a national network of 30 post acute care facilities, across the United States. The sites within this network provide skilled nursing, long term care, rehabilitation and memory care to post acute care patients. 2015 data with 15,691 encounters with Dementia patients, showed a mean performance rate of 63.93%.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
A 2007 analysis of medical records and caregiver surveys for 378 patients with dementia found that only 50% of patients received an assessment of their cognitive status in the previous 12 months.(1) Another study surveying clinicians practicing in VA medical centers found that only two thirds of clinicians reported regularly performing a standardized assessment of cognitive functioning.(2) Despite the availability and dissemination of established best practice guidelines, there is still wide variation in physician practice patterns in dementia care. The quality of currently available studies limits the ability to draw strong conclusions. (3)

1. Chodosh J, Mittman BS, Connor KI. Caring for patients with dementia: How good is the quality of care? Results from three health systems. J Am Geriatr Soc. 2007 Aug;55(8):1260-8.
2. Rosen CS, Chow HC, Greenbaum MA, et al. How well are clinicians following dementia practice guidelines? Alzheimer Dis Assoc Disord. 2002;16(1): 15-23.
3. Sivananthan SN1, Puyat JH, McGrail KM. Variations in self-reported practice of physicians providing clinical care to individuals with dementia: a systematic review. J Am Geriatr Soc. 2013 Aug;61(8):1277-85. doi: 10.1111/jgs.12368. Epub 2013 Jul 26.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
While this measure is included in several federal reporting programs, those programs have not yet made disparities data available for us to analyze and report.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
We are not aware of any publications/evidence outlining disparities specifically related to assessing cognition in dementia patients. However, a systematic review and meta-analysis of the use of dementia treatment, care, and research identified significant racial and ethnic disparities in western countries, particularly the United States. Overall, the authors found "consistent evidence, mostly from the United States, that [minority ethnic] people accessed diagnostic services later in their illness, and once they received a diagnosis, were less likely to access anti-dementia medication, research trials, and 24-hour care."

Cooper C, Tandy AR, Balamurali TB, Livingston G. A systematic review and meta-analysis of ethnic differences in use of dementia treatment, care, and research. Am J Geriatr Psychiatry. 2010 Mar;18(3):193-203.

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Dementia is a chronic condition that poses a major and growing threat to the public's health. Improving the effectiveness of care and optimizing patient outcomes will become increasingly important as the population of the United States ages.

- Today, 5.3 million Americans are living with Alzheimer's disease, including an estimated 200,000 under the age of 65. By 2050, up to 16 million will have the disease. (1)
- One in nine people age 65 and older (11 percent) has Alzheimer's disease. (1)
- About one-third of people age 85 and older (32 percent) have Alzheimer's disease. (1)
- The number of new cases of Alzheimer's increases dramatically with age: in 2015, there will be approximately 61,000 new cases among people age 65 to 74, 172,000 new cases among people age 75 to 84, and 240,000 new cases among people age 85 and older (the "oldest-old"). This translates to approximately two new cases per 1,000 people age 65 to 74, 13 new cases per 1,000 people age 75 to 84, and 39 new cases per 1,000 people age 85 and older.
- Because of the increase in the number of people over 65 in the United States, the annual incidence of Alzheimer's and other dementias is projected to double by 2050. (1)
- More than 17 percent of women and approximately 9 percent of men reaching the age of 65 would ultimately develop dementia (estimated lifetime risk). (1)
- Alzheimer's disease was the sixth-leading cause of death across all ages in the United States in 2013. It was the fifth-leading cause of death for those aged 65 and older in 2013. (1)
- People with Alzheimer's disease and other dementias have more than three times as many hospital stays as other older

people. (1)

- There are 780 hospital stays per 1,000 Medicare beneficiaries age 65 and older with Alzheimer's disease and other dementias compared with 234 hospital stays per 1,000 Medicare beneficiaries age 65 and older without these conditions. (1)
- Individuals newly diagnosed with Alzheimer's disease have higher health care use and costs in the year prior to diagnosis and in the 2 subsequent years after diagnosis than those who do not receive this diagnosis. Enrollees with a new diagnosis of Alzheimer's disease had $2,529 more in health care costs (medical and pharmacy) in the year prior to diagnosis, $10,126 more in costs in the year following diagnosis, and $6,251 more in costs in the second year following diagnosis. (1)
- Total Medicaid spending for people with Alzheimer's disease and other dementias is projected to be $41 billion in 2015 (in 2015 dollars). (1)
- The identification of high-quality dementia care guidelines and measures across settings has also been identified as a key strategy in HHS´s National Plan to Address Alzheimer´s Disease. In particular, the plan suggests that measures are needed that can track whether recommended care is being provided. These measures should be based on guidelines tailored to the stages of the disease, addressing the physical, cognitive, emotional, and behavioral symptoms of AD, and covering the myriad care settings in which care is delivered.(2)

1c.4. Citations for data demonstrating high priority provided in 1a.3

1. Alzheimer's Association. 2015 Alzheimer's Disease Facts and Figures. Alzheimer's Association ; 2015. http://www.alz.org/facts/downloads/facts_figures_2015.pdf. Accessed January 11, 2016.

2. U.S. Department of Health and Human Services. National plan to address alzheimer´s disease. Available at: http://aspe.hhs.gov/daltcp/napa. Accessed May 16, 2012.

1c.5. If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. *(Describe how and from whom their input was obtained.)*

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology, Neurology : Cognitive Impairment/Dementia

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
eCQM Library webpage at: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/eCQM_Library.html
Value set details at VSAC webpage: https://vsac.nlm.nih.gov

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is an eMeasure **Attachment:** EP_CMS149v4_NQFXXXX_DEMENTIA_Cognitive.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** EP_eCQM_DementiaCognitive_ValueSets_Jan2016.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
NOT APPLICABLE

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
Patients for whom an assessment of cognition is performed and the results reviewed at least once within a 12 month period
Definition:
Cognition can be assessed by the clinician during the patient´s clinical history. Cognition can also be assessed by direct examination of the patient using one of a number of instruments, including several originally developed and validated for screening purposes. This can also include, where appropriate, administration to a knowledgeable informant. Examples include, but are not limited to:
-Blessed Orientation-Memory-Concentration Test (BOMC)
-Montreal Cognitive Assessment (MoCA)
-St. Louis University Mental Status Examination (SLUMS)
-Mini-Mental State Examination (MMSE) [Note: The MMSE has not been well validated for non-Alzheimer´s dementias]
-Short Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE)
-Ascertain Dementia 8 (AD8) Questionnaire
-Minimum Data Set (MDS) Brief Interview of Mental Status (BIMS) [Note: Validated for use with nursing home patients only]
-Formal neuropsychological evaluation
-Mini-Cog

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*

At least once during the 12 consecutive month measurement period

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
For EHR:
HQMF eMeasure developed and is included in this submission.

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
All patients, regardless of age, with a diagnosis of dementia

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
For EHR:
HQMF eMeasure developed and is included in this submission.

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Exceptions: Documentation of medical reason(s) for not assessing cognition (eg, patient with very advanced stage dementia, other medical reason)
Documentation of patient reason(s) for not assessing cognition

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
Exceptions are used to remove a patient from the denominator of a performance measure when the patient does not receive a therapy or service AND that therapy or service would not be appropriate due to patient-specific reasons. The patient would otherwise meet the denominator criteria. Exceptions are not absolute, and are based on clinical judgment, individual patient characteristics, or patient preferences. The PCPI exception methodology uses three categories of reasons for which a patient may be removed from the denominator of an individual measure. These measure exception categories are not uniformly relevant across all measures; for each measure, there must be a clear rationale to permit an exception for a medical, patient, or system reason. Examples are provided in the measure exception language of instances that may constitute an exception and are intended to serve as a guide to clinicians. For measure Dementia: Cognitive Assessment, exceptions may include medical reason(s) (eg, patients with very advanced stage dementia, other medical reason) or patient reason(s) for not assessing cognition. Where examples of exceptions are included in the measure language, value sets for these examples are developed and included in the eMeasure. Although this methodology does not require the external reporting of more detailed exception data, the PCPI recommends that physicians document the specific reasons for exception in patients' medical records for purposes of optimal patient management and audit-readiness. The PCPI also advocates the systematic review and analysis of each physician's exceptions data to identify practice patterns and opportunities for quality improvement.

For EHR:
HQMF eMeasure developed and is included in this submission.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
Consistent with CMS' Measures Management System Blueprint and recent national recommendations put forth by the IOM and NQF to standardize the collection of race and ethnicity data, we encourage the results of this measure to be stratified by race, ethnicity, administrative sex, and payer and have included these variables as recommended data elements to be collected.

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
No risk adjustment or risk stratification

If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
No risk adjustment or risk stratification

S.15. **Detailed risk model specifications** *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.

S.15a. **Detailed risk model specifications** (*if not provided in excel or csv file at S.2b*)
No risk adjustment or risk stratification

S.16. **Type of score:**
Rate/proportion
If other:

S.17. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
To calculate performance rates:
1. Find the patients who meet the initial population (ie, the general group of patients that a set of performance measures is designed to address).
2. From the patients within the initial population criteria, find the patients who qualify for the denominator (ie, the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial population and denominator are identical.
3. From the patients within the denominator, find the patients who meet the numerator criteria (ie, the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator
4. From the patients who did not meet the numerator criteria, determine if the provider has documented that the patient meets any criteria for exception when denominator exceptions have been specified for medical reason(s) (eg, patients with very advanced stage dementia, other medical reason), or patient reason(s) for not assessing cognition. If the patient meets any exception criteria, they should be removed from the denominator for performance calculation. --Although the exception cases are removed from the denominator population for the performance calculation, the exception rate (ie, percentage with valid exceptions) should be calculated and reported along with performance rates to track variations in care and highlight possible areas of focus for QI.

If the patient does not meet the numerator and a valid exception is not present, this case represents a quality failure.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No diagram provided

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable. The measure is not based on a sample.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.

Not applicable. The measure is not based on a survey.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Patient eligibility is determined by a set of defined criteria relevant to a particular measure. If data required to determine patient eligibility are missing, those patients/cases would be ineligible for inclusion in the denominator and therefore the patient/case would be deleted.

If data required to determine if a denominator eligible patient qualifies for the numerator (or has a valid exclusion/exception) are missing, this case would represent a quality failure.

S.23. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Electronic Clinical Data, Electronic Clinical Data : Electronic Health Record

S.24. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Not applicable

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Group/Practice, Clinician : Individual, Clinician : Team

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Ambulatory Care : Clinician Office/Clinic, Ambulatory Care : Urgent Care, Behavioral Health/Psychiatric : Inpatient, Other, Post Acute/Long Term Care Facility : Nursing Home/Skilled Nursing Facility
If other: Occupational Therapy Services, 'Domiciliary', Rest Home or Custodial Care Services

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
Not applicable. The measure is not a composite.

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
testing_form_for_trial_use_dementia_final.docx

NATIONAL QUALITY FORUM
Measure Testing Form for Trial Approval Program

Measure Title: Dementia – Cognitive Assessment
Date of Submission: 01/15/2016
Type of Measure:

☐ Composite	☐ Outcome (*including PRO-PM*)
☐ Cost/resource	☒ Process
☐ Efficiency	☐ Structure

Instructions

- A measure submision that is to be considered for the Trial Approval Program must complete this form in its entirety. Either a test data set provided by the measure developer, or the use of the Bonnie tool is acceptable to provide prelminary testing results,
- **For all measures being submitted for potenital acceptance into the Trial Approval Program, each section must be filled out as completely as possible.**
- Respond to all questions as instructed with answers immediately following the question. All information on testing of either a sample data set or results from Bonnie testing that can demonstrate, to the extent possible, the the measure meets the reliability and validity must be in this form..
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions here.

DATA and SAMPLING INFORMATION

1. DATA/SAMPLE USED FOR PRELMINARY TESTING OF THIS MEASURE

It is important that the measure developer use a data set to conduct preliminary testing in order to evauate the measure logic and the inclusions/exclusions for the population used in the measure.

1.1. What type of data was used for testing? (*The measure developer must provide a test data set that will provide some initial information to be used for the evaluation, or it can use the Bonnie testing tool to create a sample data set using synthesized patients.)* Please indicate whether the test data set used was provided through the measure developer, or through the Bonnie tool.

A synthetic patient deck was created, using the Bonnie tool.

1.2. If Bonnie was NOT used, please identify the specifications for the test dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured)* Not applicable

1.3. What levels of analysis were tested (either through the test data set or Bonnie)? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*) *in order to determine its suitability for inclusion into the Trial Approval Program.,*

Measure Specified to Measure Performance of:	Measure Tested at Level of:

☒ individual clinician	☒individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ other: Click here to describe	☐ other: Click here to describe

1.4. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis)*

32 patients, with a diagnosis of Dementia & Mental Degenerations, were included in the Bonnie test deck. 3 patients had a diagnosis of Severe Dementia. Tables summarizing age, race and ethnicity, and sex are included below:

Age	**Count**
0-17	1
18-75	25
75+	6

Race & Ethnicity	**Count**
Not Hispanic or Latino	**32**
American Indian or Alaska Native	2
Other	30
Hispanic or Latino	**0**

Sex	**Count**
F	17
M	15

1.5. Please fill out the testing spreadsheet found at this link. The spreadsheet must be completed in its entirety, to the extent possible, in order to provide a basis for evaluation to determine the acceptability of the measure for inclusion in the Trial Approval program. *Any questions regarding the completion of this form can be directed to NQF Staff here.*

A spreadsheet was attached to the submission form.

<u>RELIABILITY AND VALIDITY ASSESSMENTS</u>

Note: The information provided in this next section is intended to aid the Standing Committee and other stakeholders in understanding to what degree the measure is both reliable and valid. While it is not possible to provide comprehensive results due to the lack of actual testing data, the developer needs to provide as much information as possible based on their interpretation of the results from the sample test data.

2.1 Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score. **What is your interpretation of the results in terms of demonstrating reliability**? (i.*e., what do the sample results mean and what are the norms for the test conducted?*) Please summarize the plan for future testing of reliability if the measure is accepted into the Trial Approval Program. Include descriptions of:

- Inter-abstractor reliability, and data element reliability of all critical data elements
- Computation of the performance measure score (e.g., signal-to-noise analysis)?

Reliability of the computed measure score will be measured as the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in physician performance. Reliability at the level of the specific physician is given by:

Reliability = Variance (physician-to-physician) / [Variance (physician-to-physician) + Variance (physician-specific-error]

Reliability is the ratio of the physician-to-physician variance divided by the sum of the physician-to-physician variance plus the error variance specific to a physician. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in physician performance.

Reliability testing will be performed by using a beta-binomial model. The beta-binomial model assumes the physician performance score is a binomial random variable conditional on the physician's true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates.

Reliability will be estimated at two different points, at the minimum number of quality reporting events for the measure and at the mean number of quality reporting events per physician.

2.2 Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score. **What is your interpretation of the results in terms of demonstrating validity**? (i.*e., what do the results mean and what are the norms for the test conducted?*). Please summarize the plan for future testing of validity if the measure is accepted into the Trial Approval Program. Include the method(s) of validity testing and what it will test (describe the steps—do not just name a method; what will be tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis will be used used)

Face validity of the measure score as an indicator of quality will be systematically assessed as follows.

The expert panel will be asked to rate their agreement with the following statement:

The scores obtained from the measure as specified will provide an accurate reflection of quality and can be used to distinguish good and poor quality.

Scale 1-5, where 1= Strongly Disagree; 3= Neither Agree nor Disagree; 5= Strongly Agree

2.3 Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion. **What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results?** (*i.e., the value outweighs the burden of increased data collection and analysis)*. Please summarize the plan for future testing of exclusions if the measure is accepted into the Trial Approval Program. Describe the method of testing exclusions and what it will test (describe the steps—do not just name a method; what will be tested, e.g., whether exclusions affect overall performance scores; what statistical analysis will be used)

Measure exceptions will be analyzed for frequency and variability across providers.

2.4 Risk Stratification (applicable ONLY to outcome or resource use measures). If an outcome or resource use measure will not be risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities. If risk adjustment/stratification is needed then please describe the conceptual/clinical and statistical methods and criteria that will be used to select patient factors (clinical factors or sociodemographic factors) that will be used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Not applicable

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic health records (EHRs)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment **Attachment:** Dementia_Cognitive_Assessment_Feasibility_testing_attachment.pdf

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

The feasibility of this measure was assessed by two entities, as outlined in the feasibility attachment. The measure exceptions, for this measure, are not currently being captured in a structured format within the two EHR products assessed for feasibility. Exception information is extracted from documentation within the medical record, explaining why the patient did not receive the standard of care. Although this information is currently entered as text, it is identifiable in the medical record. Future iterations of vendor products will likely allow health care providers to collect exceptions information in a structured format. The measure is feasible for implementation and data collection in EHRs, as specified, in its current format. No modifications have been made to this measure, due to issues with data collection, sampling or cost.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

The Measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, eg, use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Measures require a license agreement between the user and the AMA, (on behalf of the PCPI). Neither the AMA, PCPI nor its members shall be responsible for any use of the Measures.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting	Payment Program Meaningful Use Stage 2 http://www.cms.gov/Regulationsand-Guidance/Legislation/EHRIncentivePrograms/Stage_2.html http://www.cms.gov/Medicare/Quality- Initiatives-Patient-Assessment-Instruments/pqrs/index.html PQRS Professional Certification or Recognition Program NeuroPI https://tools.aan.com/practice/pip/

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

Physician Quality Reporting System (PQRS)-Sponsored by the Centers for Medicare and Medicaid Services (CMS)
Purpose: PQRS is a national reporting program that uses a combination of incentive payments and payment adjustments to promote reporting of quality information by eligible professionals (EPs). The program provides an incentive payment to practices with EPs (identified on claims by their individual National Provider Identifier [NPI] and Tax Identification Number [TIN]). Eps satisfactorily report data on quality measures for covered Physician Fee Schedule (PFS) services furnished to Medicare Part B Fee-for-Service (FFS) beneficiaries (including Railroad Retirement Board and Medicare Secondary Payer). Beginning in 2015, the program also applies a payment adjustment to EPs who do not satisfactorily report data on quality measures for covered professional services.

Dementia Cognitive Assessment is included in PQRS as part of the Dementia Measure Set- data is not available for the individual measure. It is our understanding that CMS is also planning to move towards publicly reporting physician data via Physician Compare.

Meaningful Use Stage 2 (EHR Incentive Program) – Sponsored by the Centers for Medicare and Medicaid Services (CMS)
The Medicare and Medicaid EHR Incentive Programs provide incentive payments to eligible professionals, eligible hospitals, and critical access hospitals (CAHs) as they adopt, implement, upgrade or demonstrate meaningful use of certified EHR technology. These professionals are eligible for incentive payments for the "meaningful use" of certified EHR technology, if all program requirements are met, including successful implementation and reporting of program measures, which include this measure, to demonstrate meaningful use of EHR technology.

The measure is currently utilized in NeuroPI. This measure is included in a Maintenance of Certification module addressing needs for patients with dementia. There have been two editions of the dementia module released. Performance data on specific measures is not available, as it is not stored or reconciled in the system.

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. **Progress on Improvement. (Not required for initial endorsement unless available.)**
Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

NOT APPLICABLE

4b.2. **If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.**
NOT APPLICABLE

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We are not aware of any unintended consequences at this time, but we take unintended consequences very seriously and therefore continuously monitor to identify actions that can be taken to mitigate them.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment:

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** PCPI
Co.2 **Point of Contact:** Caryn, Davidson, caryn.davidson@ama-assn.org, 312-464-4465-
Co.3 **Measure Developer if different from Measure Steward:** PCPI
Co.4 **Point of Contact:** Caryn, Davidson, pcpimeasures@ama-assn.org

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
PCPI measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study are invited to participate as equal contributors to the measure development process. In addition, the PCPI strives to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.

Jerry C. Johnson, MD (Co-Chair) (geriatric medicine)
Germaine Odenheimer, MD (Co-Chair) (neurology)
François Boller, MD, PhD, FAAN (neurology)
Soo Borson, MD (geriatric psychiatry)
Charles A. Cefalu, MD, MS (geriatric medicine)

Mirean Coleman, MSW, LICSW, CT (social work)
Patricia C. Davis, MD, MBA, FACR (radiology)
Mary Ann Forciea, MD (internal/geriatric medicine)
Elizabeth M. Galik, PhD, CRNP (nursing)
Laura N. Gitlin, PhD (occupational therapy)
Helen H. Kyomen, MD, MS (geriatric and adult psychiatry)
Katie Maslow, MSW (patient advocacy representative)
Haydee Muse, MD (health plan representative)
Bruce E. Robinson, MD, MPH (geriatric medicine)
Robert Paul Roca, MD, MPH, MBA (geriatric psychiatry)
Amy E. Sanders, MD (geriatric neurology)
Jason E. Schillerstrom, MD (geriatric psychiatry)
Joseph W. Shega, MD (geriatric medicine, hospice and palliative medicine)
Eric G. Tangalos, MD, FACP, AGSF, CMD (internal/geriatric medicine)
Joan M. Teno, MD, MS (internal medicine)
Brian K. Unwin, MD, FAAFP (family medicine)
John Robert Absher, MD (neurology) -- Liaison to American Academy of Neurology's Quality Measurement and Reporting Subcommittee

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2011

Ad.3 **Month and Year of most recent revision:** 08, 2015

Ad.4 **What is your frequency for review/update of this measure?** Supporting guidelines for this measure are reviewed on an annual basis. Specification and coding upd

Ad.5 **When is the next scheduled review/update for this measure?** 12, 2016

Ad.7 **Disclaimers:** Physician Performance Measures (Measures) and related data specifications have been developed by the American Medical Association (AMA)-convened Physician Consortium for Performance Improvement® (PCPI®). These Measures are not clinical guidelines and do not establish a standard of medical care, and have not been tested for all potential applications.

The Measures, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, eg, use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the Measures require a license agreement between the user and the AMA, (on behalf of the PCPI). Neither the AMA, PCPI nor its members shall be responsible for any use of the Measures.

THE MEASURES AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Limited proprietary coding is contained in the Measure specifications for convenience. Users of the proprietary code sets should obtain all necessary licenses from the owners of these code sets. The AMA, the PCPI and its members disclaim all liability for use or accuracy of any Current Procedural Terminology (CPT®) or other coding contained in the specifications. Use of CPT coding beyond fair use requires a license from the AMA. CPT® contained in the Measure specifications is copyright 2004-2014 American Medical Association. LOINC®copyright 2004-2014 Regenstrief Institute, Inc. This material contains SNOMED Clinical Terms® (SNOMED CT®) copyright 2004-2014 International Health Terminology Standards Development Organisation. ICD-10 copyright 2014 World Health Organization. All Rights Reserved.

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2876
Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute ischemic stroke hospitalization with claims-based risk adjustment for stroke severity
Measure Steward: Centers for Medicare & Medicaid Services (CMS)
Brief Description of Measure: This stroke mortality measure estimates the hospital-level, risk-standardized mortality rate (RSMR) for patients discharged from the hospital with a principal discharge diagnosis of acute ischemic stroke. The outcome is all-cause 30-day mortality, defined as death from any cause within 30 days of the index admission date, including in-hospital death, for stroke patients. This is a newly developed measure with a cohort and outcome that is harmonized with the CMS's current publicly reported claims-based stroke mortality measure and includes the National Institutes of Health (NIH) Stroke Scale as an assessment of stroke severity in the risk-adjustment model. This measure uses Medicare fee-for-service (FFS) administrative claims for the cohort derivation, outcome, and risk adjustment.
Developer Rationale: Stroke is the fifth most common cause of death, affecting approximately 795,000 people in the United States annually, and has a mortality rate of 17% [Go et al., 2014; Kochanek et al., 2014]. Stroke is also a leading cause of disability in the United States, which can lead to increased dependency on the health care system and higher subsequent costs associated with this care [Centers for Disease Control and Prevention, 2005]. Mortality following stroke – an important adverse outcome that can be measured reliably and objectively, and that is influenced by the quality of care provided to patients during their initial hospitalization – is an appropriate measure of quality of care [DesHarnais et al., 1988; Weir et al, 2001]. Specifically, post-stroke mortality rates have been shown to be influenced by critical aspects of care such as response to complications, speediness of delivery of care, organization of care, and appropriate imaging [Smith et al., 2006; Reeves et al., 2009; Lingsma et al., 2008; Hong et al., 2008]. This work demonstrates the relationship between hospital organizational factors and performance on the stroke mortality measure and supports the ability of hospitals to impact these rates.

The goal of outcome measurement is to identify institutions whose performance is better or worse than would be expected based on their patient case mix by risk-adjusting for patients' conditions and stroke severity at the time of hospital admission. The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level RSMRs following hospitalization for acute ischemic stroke. Measurement of patient mortality allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures.

Rationale for Development of an Updated Claims-Only Stroke Mortality Measure
Current outcome measures use administrative claims data from the year prior to the index admission in the risk adjustment models. Stakeholders, including the AHA/ASA and other professional organizations, have highlighted the importance of including stroke severity in mortality measures for risk adjustment. Several studies have demonstrated that initial stroke severity is the strongest predictor of mortality in ischemic stroke patients [Smith et al., 2010; Nedeltchev et al., 2010; Fonarow et al., 2012].
This new claims-based stroke mortality measure addresses these stakeholder preferences and improves model performance by updating the current publicly reported claims-based stroke mortality measure to incorporate stroke severity scores into the risk-adjustment model. Advancements in clinical practice to incorporate new clinical assessments

in administrative coding systems have made it possible to integrate these data into measures of hospital performance. The NIH Stroke Scale, which was created in 1989 and is widely used in routine stroke care, is collected in the GWTG-Stroke Registry, which has over 1,700 hospitals throughout the U.S. [Fonarow et al., 2014]. The NIH Stroke Scale is a 15-item neurologic examination stroke scale used to provide a quantitative assessment of stroke related neurologic deficit, by evaluating the effect of acute ischemic stroke on the levels of consciousness, language, neglect, visual-field loss, extraocular movement, motor strength, ataxia, dysarthria, and sensory loss. The NIH Stroke Scale is designed to be a simple, valid, and reliable tool that can be administered at the bedside consistently by physicians, nurses, or therapists. The use of the NIH Stroke Scale to assess stroke severity upon acute ischemic stroke patient presentation is recommended in the AHA/ASA Class I guidelines. Furthermore, the NIH Stroke Scale scores will be coded in the ICD-10-CM coding system beginning in October 2016, allowing it to be used in this measure. Inclusion of stroke severity data will not only address stakeholder preferences, but may also improve the discrimination of the risk models.

References:

1. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. Circulation. Jan 21 2014;129(3):e28-e292.
2. Kochanek KD MS, Xu JQ, Arias E. Mortality in the United States, 2013. NCHS data brief, no 178. 2014.
3. Centers for Disease Control and Prevention (CD). Prevalence and most common causes of disability among adults: United States, 2005. MMWR Morb Mortal Wkly Rep. 2009;58:421-426.
4. Casper ML NI, Croft JB, Nilasena DS. Atlas of Stroke Hospitalizations Among Medicare Beneficiaries. 2008.
5. DesHarnais SI, Chesney JD, Wroblewski RT, Fleming ST, McMahon LF, Jr. The Risk-Adjusted Mortality Index. A new measure of hospital performance. Med Care. Dec 1988;26(12):1129-1148.
6. Weir NU, Sandercock PA, Lewis SC, Signorini DF, Warlow CP. Variations between countries in outcome after stroke in the International Stroke Trial (IST). Stroke. Jun 2001;32(6):1370-1377.
7. Smith MA, Liou JI, Frytak JR, Finch MD. 30-day survival and rehospitalization for stroke patients according to physician specialty. Cerebrovascular diseases (Basel, Switzerland). 2006;22(1):21-26.
8. Reeves MJ, Smith E, Fonarow G, Hernandez A, Pan W, Schwamm LH. Off-hour admission and in-hospital stroke case fatality in the get with the guidelines-stroke program. Stroke. Feb 2009;40(2):569-576.
9. Lingsma HF, Dippel DW, Hoeks SE, et al. Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey. Journal of neurology, neurosurgery, and psychiatry. Aug 2008;79(8):888-894.
10. Hong KS, Kang DW, Koo JS, et al. Impact of neurological and medical complications on 3-month outcomes in acute ischaemic stroke. European journal of neurology : the official journal of the European Federation of Neurological Societies. Dec 2008;15(12):1324-1331.
11.Smith EE, Shobha N, Dai D, et al. Risk score for in-hospital ischemic stroke mortality derived and validated within the Get With the Guidelines-Stroke Program. Circulation. 2010;122(15):1496-1504.
12. Nedeltchev K, Renz N, Karameshev A, et al. Predictors of early mortality after acute ischaemic stroke. Swiss Medical Weekly. 2010;140(17-18):254-259.
13. Fonarow GC, Saver JL, Smith EE, et al. Relationship of national institutes of health stroke scale to 30-day mortality in medicare beneficiaries with acute ischemic stroke. J Am Heart Assoc. Feb 2012;1(1):42-50.
14. Adams HP, Jr., Davis PH, Leira EC, et al. Baseline NIH Stroke Scale score strongly predicts outcome after stroke: A report of the Trial of Org 10172 in Acute Stroke Treatment (TOAST).
15. Fonarow GC, Alberts MJ, Broderick JP, et al. Stroke outcomes measures must be appropriately risk adjusted to ensure quality care of patients: a presidential advisory from the American Heart Association/American Stroke Association. Stroke. May 2014;45(5):1589-1601.

Numerator Statement: The outcome for this measure is 30-day, all-cause mortality. We define mortality as death from any cause within 30 days of the index admission for patients with a principal discharge diagnosis of acute ischemic stroke.

Denominator Statement: The cohort includes inpatient admissions to all non-federal, short-term, acute care hospitals for Medicare FFS patients age 65 years and older with a principal discharge diagnosis of acute ischemic stroke.

Additional details are provided in S.9 Denominator Details.

Denominator Exclusions: The measure excludes admissions for patients:
1. With inconsistent or unknown vital status or other unreliable data;
2. Enrolled in the Medicare hospice program at any time in the 12 months prior to the index admission, including the first day of the index admission; and
3. Discharged against medical advice (AMA).
For patients with more than one admission for stroke in a given year, only one index admission for that condition is randomly selected for inclusion in the cohort.

Measure Type: Outcome
Data Source: Administrative claims, Electronic Clinical Data : Registry, Other
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- This is a newly-developed claims-based measure that calculates hospitals' 30-day risk-standardized mortality rate for patients who have been hospitalized with an ischemic stroke. This measure is harmonized with the CMS's current publicly reported claims-based stroke mortality measure and includes the National Institutes of Health (NIH) Stroke Scale as an assessment of stroke severity in the risk-adjustment model. In 2012, the Neurology Steering Committee considered a similar measure that did not include the NIHSS in the risk-adjustment approach but the Committee could not come to consensus about the measure and it was withdrawn from consideration by the developer.
- As a rationale for measuring this health outcome, the developer states that many hospital processes have been associated with lower stroke mortality rates within 30 days of hospital admission including prevention of, and response to, complications, speediness of delivery of care, organization of care, appropriate imaging, patient safety, and coordinated transitions to the outpatient environment.
- The developer reports studies demonstrate appropriate, guideline-recommended care and timely treatment for stroke patients can reduce the risk of mortality within 30 days of hospital admission.

Question for the Committee:

- *Does the SC agree that at least one hospital process identified by the developer impacts ischemic stroke mortality rates within 30 days of admission?*

Guidance from the Evidence Algorithm : Health outcome (Box 1) →relationship between outcome and at least one healthcare action identified/supported by stated rationale (Box 2) → Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

Per the developer, they combined two data sources (for model development purposes only): July 2011 – June 2014 Medicare Administrative claims and 2013 AHA/ASA GWTG-Stroke Registry. Values for the NIHSS were obtained from the registry as a surrogate for NIHSS scores that will be obtained from ICD-10 codes beginning in October 2016.

- **Sample size**: 188,975 patients at 1,511 hospitals
- **Mean** risk-standardized mortality rate (RSMR): 14.53%
- **Median** risk-standardized mortality rate (RSMR): 14.48%
- **Range**: 10.75% to 18.98%
- **Interquartile range**: 13.52% to 15.56%

Developer also provided the results of CMS's current publicly reported claims-based stroke mortality measure (does not include the NIH Stroke Scale in the risk-adjustment model) as reported in 2014 update to Hospital Compare which demonstrated variation across hospitals.

- **Sample size:** 520,111 admissions from 4,506 hospitals
- **Median** risk-standardized mortality rate (RSMR): 15.3%
- **Interquartile range**: 8.6% to 23.8%

Disparities

- The developer provides the following information (July 2011-June 2014):

	#hospitals	*# admissions*	*Minimum rate*	*Median rate*	*Maximum rate*
Dual eligibles					
Low proportion	551	19,859	11.27	14.25	17.37
High proportion	294	14,376	11.34	14.22	17.40
African Americans					
Low proportion	492	13,497	12.15	14.27	17.37
High proportion	294	17,429	11.27	14.15	17.15
AHRQ SES score					
Low proportion	295	11,563	11.77	14.27	17.15
High proportion	294	22,187	11.34	14.17	16.59

Questions for the Committee:

- *Does this gap in care warrant a national performance measure?*
- *Are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **The evidence supports the idea that hospital processes can influence risk-adjusted mortality rates.

**This is a tricky problem. As a starting point, mortality has not historically been considered the gold-standard outcome for stroke trials, unlike trials in other acute conditions. While there are many reasons for this, one important reason is that mortality is not always an outcome that is important to patients when facing a severe stroke. Many patients, when facing a trade off between survival with severe disability and death, would choose death.

There is face validity and some (relatively limited) data supporting the notion that mortality is associated with high quality processes of care and/or structural variables. These data, though often use very crude proxies for high structural variables that may correlate with a variety of factors that influence mortality. For example, while it may be the case that Get with the Guidelines Stroke hospitals do indeed have cultures of quality that lead to good outcomes, it could also be the case that they attract different patient populations, participate in GWTG because of their patient preferences or have other unique factors that influence mortality other than quality of care (i.e. intensity of care, access to technology, etc.) It is also the case that some quality factors (i.e. stroke volume) are correlate relatively poorly with mortality.

This become potentially problematic as mortality has also been associated with factors that are not obviously quality related such as high intensity care (i.e. hemicraniectomy) and preferences for high intensity care (i.e. lack of DNR orders).

There are little data to inform whether quality-based mortality associations or non-quality-based associations are larger. In general, the associations between quality factors and mortality are fairly weak. It is also entirely possible that high quality processes and intensity of care are also correlated. It may well be that the "quality" effects are "intensity" effects or vice versa. Its rather a leap of faith to say that an adjustment mortality measure is measuring "quality" as opposed to "intensity" or "unmeasured preferences."

1b. Performance Gap

Comments: **The evidence presented does not support the idea that there is a substantial gap in care between hospitals with low/high proportions of duals or African Americans or people with lower SES. The rates appear to be nearly identical.

**Variation in mortality exists at the hospital level and exists independently of measured patient factors. The extent to which this difference is do to unmeasured patient factors that correlate at the hospital level is uncertain. In addition, the extent to which this variation can be reduced is also uncertain. For acute stroke, there are very limited interventions (that are already in wide use) that reduce short-term mortality.

The sponsors' data does not appear to identify racial or SES disparities in risk standardized mortality at the hospital level.

1c. High Priority (previously referred to as High Impact)

Comments: **No comments provided

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): The developer listed administrative claims, electronic clinical data, and registry data as the data sources but this measure is specified for administrative claims only.

Specifications:

- The measure is specified as a facility-level measure for the hospital/acute care setting.
- The denominator includes all Medicare FFS beneficiaries, age 65 and over, with a principal discharge diagnosis of acute ischemic stroke.
    - In the 2012 evaluation of the stroke mortality measure, in an effort to harmonize with other mortality/readmissions measures, the developer included all-payer patients ages 18 and over (rather than Medicare FFS patient ages 65+ only). Other mortality measures can be used in either of two patient cohorts: (1) patients 65 years or older, or (2) patients 18 years or older.
- ICD-9 and ICD-10 codes are provided; ICD-9 to ICD-10 crosswalk available in the data dictionary found in S.2b.
- Exclusions include inconsistent or unknown vital status, enrolled in Medicare hospice any time in the 12 months prior to the index admission including first day of admission, and discharged AMA; exclusion details provided.
- As noted earlier, the NIHSS values will be available via ICD-10 dx code beginning October 2016.
- This outcome measure is risk-adjusted using a statistical risk-adjustment model with 20 factors including age, ED transfer, NIH Stroke Scale score, and multiple comorbidities.
- The calculation algorithm, included in S.18, describes how the risk-standardized mortality ratio is calculated.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high

proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☒ **No**

Method(s) of reliability testing:

- The dataset used for testing included Medicare Parts A and B claims, Medicare Enrollment Database(EDB), and AHA/ASA Get With The Guidelines (GWTG) Stroke Registry. The registry data were used as a surrogate for data that will eventually come from claims once ICD-10 codes for stroke severity are available and consistently used by hospitals. A subset of 1,520 non-federal, acute inpatient US hospitals that participate in GWTG Stroke Registry were included in the dataset.
    - For measure development purposes only, the developers linked two data sources: Medicare claims and GWTG Stroke Registry data. Registry data were used to get the NIH Stroke Scale score - ICD-10 codes for the NIH Stroke Scale are scheduled to be released in October 2016; therefore, available via claims data.
    - The developers then merged Medicare Part A Inpatient, Outpatient, and Part B Outpatient claims with the GWTG Stroke Registry stroke severity data from July 1, 2011 to June 30, 2014 which resulted in the following data sample (**Dataset 1**):
        - Number of hospitals: 1,511
        - Number of total admissions: 188,975
        - First half of split sample (development sample): 94,466 admissions; 1,473 hospitals
        - Second half of split sample (validation sample): 94,509 admissions; 1,462 hospitals
- Although the developer reported assessing data element reliability by comparing model variable frequencies and odds ratios from logistic regression models across the most recent three years of data, NQF does not consider temporal consistency to be a valid method of demonstrating reliability of data elements.

- Developers used a split-sample (or "test-retest") methodology to test the measure score reliability; this is an appropriate method. Developers randomly assigned half of the patients in each hospital to two separate groups, calculated the performance measure score for each hospital in each of the two groups, and compared the agreement between each hospital's paired scores using the intra-class-correlation coefficient (ICC) and applying a correction factor to account for the overall sample size. The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospitals.

Results of reliability testing :

- Measure score reliability results:
    - The ICC values from the split-sample analysis was 0.55, indicating 51% of the variance in scores are due to differences between hospitals. According to the Landis and Koch classification, an ICC value of 51% can be interpreted as moderate agreement. However, a value of 0.7 is often regarded as a minimum acceptable reliability value.
    - The developer notes that this analysis was limited to hospitals with 12 or more cases in each split sample although the measure is not specified to include a minimum data sample of 12 cases. The developer also notes that the ICC is based on a split sample of three years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data – the measure is intended to be reported with three years of data.

Questions for the Committee:

- *Do the testing results presented by the developer demonstrate an adequate level of reliability?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *assessments of performance score reliability often examine the ability of the measure to*

differentiate between measured entities. Do the reliability testing results reported by the developer demonstrate that meaningful differences in performance can be identified?

Guidance from the Reliability Algorithm: Precise specifications (Box 1) →empiric reliability testing (Box 2) →performance score testing (Box 4) →appropriate method of testing (Box 5) → moderate certainty of reliability (box 6b)

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

- This measure estimates 30-day all-cause mortality for patients hospitalized with ischemic stroke using a risk standardized mortality ratio (RSMR), which is calculated as the ratio of the number of "predicted" to the number of "expected" deaths, multiplied by the national unadjusted mortality rate.
- As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence mortality rates through a broad range of clinical activities, including prevention of, and response to, complications, speediness of delivery of care, organization of care, appropriate imaging, patient safety, and coordinated transitions to the outpatient environment.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☒ **Measure score** ☐ **Data element testing against a gold standard** ☐ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- The developer states that face validity was systematically assessed by outside experts and the public throughout the development of this measure. The outside experts formed a working group which consisted of various expertise to ensure the measure is meaningful, useful, and well-designed. The developers also solicited public comments through a 30-day public comment period. NQF guidance requires that face validity of the measure score must explicitly address whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. It does not appear that the face validity assessment conducted by the developer conforms to NQF's requirements.
- The developer conducted empirical validity testing of the measure score by comparing this measure to a similar stroke mortality measure that uses data from the GWTG Stroke Registry (as opposed to using data from administrative claims). This registry-derived model includes a total of 188,975 hospital admissions **(see Dataset 1).** Developers compared the two measures by comparing the performance of the risk models and by correlating the RSMRs from the two models, after weighting by hospital volume.
  - The developer noted that both the registry model and the claims-based model used the same inclusion/exclusion criteria and risk-adjustment strategy **except** that the registry model used chart-based risk adjusters, such as blood pressure, which were not available in the claims data.

Validity testing results:

- The performance of the administrative claims and registry models was similar; the areas under the receiver operating characteristic (ROC) curve (or *c*-statistics) were **0.8120** and **0.7939**, respectively.
    - A c-statistic is a model discrimination statistic. A c-statistic of 0.8120 means that for 81.2% of all possible pairs of patients—one who died and one who lived—the model correctly assigned a higher probability to those who died. Generally, a c-statistic of at least 0.70 is considered acceptable. The similar c-statistics indicate that both models have similar discriminatory ability.
- The correlation coefficient between the RSMRs from the administrative claims model and the registry model was **0.96**.
- The developer states that the empirical validity testing results between the administrative claims model and the registry model proved to be similar and the *c*-statistics were "nearly identical". The developer also concluded that based on the correlation coefficient (**0.96**), the resulting measure from the administrative claims model is as good as the one from the registry model.

Questions for the Committee:

- *In the absence of NIHSS values from ICD-10 codes, is use of values from the GWTG-Stroke registry a reasonable surrogate?*
- *Do the methods and results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- To determine impact of exclusions on the cohort (188,975 hospital admissions), the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion:
    - Inconsistent or unknown vital status or other unreliable data – **0.0%**
    - Enrolled in the Medicare hospice program at any time in the 12 months prior to the index admission, including the first day of the index admission – **0.78%**
    - Discharged against medical advice (AMA) – **0.22%**
- Exclusions 1 and 2 are necessary to calculate the measure; exclusion 3 is needed for acceptability of the measure to hospitals who do not have the opportunity to deliver full care and prepare the patient for discharge.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

Description of the model

- This measure is risk-adjusted using hierarchical logistic regression model with **20 factors** to create a hospital-level 30-day risk-standardized mortality rate that simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals.
    - The model adjusts for case differences based on age, transfer from another ED, NIH Stroke Scale score, and clinical status of the patient at the time of admission (the latter using condition categories (CCs)).
    - Only comorbidities that conveyed information about the patient at the time of admission or in the 12-months prior, and not complications that arose during the course of the hospitalization, were included

in the risk-adjustment.

- To select candidate variables for this measure the developers began with the list of 42 variables included in the current publicly reported Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure and added the NIH Stroke Scale score, which was pulled from the GWTG Registry dataset. The final risk-adjustment model included 20 variables that demonstrated a relatively strong association with mortality and were clinically relevant.

Performance of the model

Discrimination statistics:

- The *c*-statistic reflects how accurately a statistical model is able to distinguish between a patient with an outcome and a patient without an outcome. C-statistic values can range from 0.5 to 1.0. A value of 0.5 indicates that the model is no better than chance at making a prediction of patients with and without the outcome of interest and a value of 1.0 indicates that the model perfectly identifies those with and without the outcome of interest . Generally, a c-statistic of at least 0.70 is considered acceptable.
- The *c*-statistic value was computed using data from **Dataset 1** which was randomly split into two samples:
  - The *c*-statistic for the 1st half of the randomly split sample (development sample) was **0.8124**.
  - The *c*-statistic for the 2nd half of the randomly split sample (validation sample) was **0.8165**.
- A *c*-statistic of **0.8124** means that for 81.2% of all possible pairs of patients—one who died and one who lived—the model correctly assigned a higher probability to those who died. The similar c-statistic in the validation model shows that the risk model has similar discriminatory ability in a dataset different from the one from which it was developed.
- Developers also noted that the model indicated a wide range between the lowest decile and highest decile which indicated the ability to distinguish between high-risk patients and low-risk patients.
  - 1st half of randomly split sample (development sample): **1.33%, 50.0%**
  - 2nd half of randomly split sample (validation sample): **1.26%, 51.57%**

Calibration statistics:

- The developers noted that the risk-decile plot based on the development dataset demonstrated excellent discrimination and good predictive ability of the risk-adjustment model because the observed values are relatively similar to the predicted values across the risk-deciles. The developers also noted that the plot for the validation dataset demonstrated similar results.
- According to the developers, the calibration values of almost 0 and almost 1, for the development sample and the validation sample, indicated good calibration of the model (i.e., the model is not overfit).

Conceptual basis and empirical support for potential inclusion of SDS factors in risk-adjustment approach

- The developer noted that although some recent literature has evaluated the relationship between patient SES or race and mortality, few studies directly address causal pathways or examine the role of the hospital in these pathways. Additionally, there is no clear consensus in the literature on which risk factors demonstrate the strongest relationship with mortality. They note that SES factors that have been examined in the mortality literature include: patient-level self-reported or documented race or ethnicity, income, and education level; occupational level; median household income; Agency for Healthcare Research and Quality (AHRQ)-validated SES index score; and the proportion of Medicaid patients served in the hospital.
  The developer identified several potential conceptual pathways to consider:
  - Relationship of socioeconomic status (SES) factors or race to health at admission.
  - Use of low-quality hospitals.
  - Differential care within a hospital.
  - Influence of socioeconomic status (SES) on mortality risk outside of hospital quality and health status.
- Based on their interpretation of the literature and analysis of the above pathways, the developers identified 3 potential SDS variables for potential inclusion in the risk-adjustment model:
  - African American race (as compared to all others)
  - Dual eligible status
  - AHRQ SES index score (based on 5-digit ZIP code data; includes percentage of people in the labor force

who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th-grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room)

- Analyses indicate that the prevalence of these 3 SDS factors varies across hospitals and are associated with the measured outcome. Of the 3 SDS factors, only African American race was significantly associated with a lower risk of mortality (OR=0.62, which indicates a protective effect). The developers note that the c-statistics for the models that included the individual SDS factors along with the original variables were similar (original=0.8176; with dual eligible=0.8176; with race= 0.8184; with AHRQ SES index=0.8176). Addition of these factors individually resulted in very little change in the RSMRs (e.g., for race, the average absolute change in hospitals' RSMRs was -0.00064%).
- Based on these results, the developer decided **NOT** to include any of the 3 SDS factors analyzed in the final risk-adjustment model.

Questions for the Committee:

- *Does the risk model adequately control for differences in case mix across providers?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*
- *Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developers provided statistical results from the dataset used to develop this measure (Dataset 1, July 2011-July 2014) which demonstrated substantial variation in RSMRs among hospitals. The median hospital RSMR was 14.29% with a range of 11.27% to 17.54%. The interquartile range was 13.63% - 15.05%.
- The developers provided additional mortality rates from the literature.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- Not Applicable

2b7. Missing Data

- The developers assessed the frequency of missing NIH Stroke Scale scores from the GWTG Stroke Registry during the development of this measure. The registry data were used as a surrogate for data that will eventually come from claims once ICD-10 codes for stroke severity are available and consistently used by hospitals. The analysis performed on the development dataset found:
  - July 1, 2011 – June 30, 2012: NIH Stroke Scale scores available in **70.78%** admissions for ischemic stroke
  - July 1, 2013 – June 30, 2014: NIH Stroke Scale scores available in **82.39%** of ischemic stroke admissions
- The developers used multiple imputation to substitute missing values of stroke severity with predicted values and found that the proportion of patients missing NIH Stroke Scale scores had little impact on hospital 30-day mortality rates. However, these analyses will need to be repeated in the future using a claims dataset.
- Due to the importance of this measure in national reporting programs and the importance of stroke severity to this measure's risk model the developer states that CMS will determine the plan for handling missing data during the measure implementation.

Guidance from Validity algorithm: Precise specifications (Box 1) → potential threats to validity assessed (Box 2) → empirical validity testing (Box 3) →measure score testing (Box 6) → method conceptually sound conceptualization (Box

7) → moderate confidence that scores are a valid indicator of quality (Box 8b)

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **The specifications appear to be consistent with the evidence.

**I'm not sure if this point refers to validity or reliability of specification however one limitation is that it is unclear if NIHSS will be reliably measured via the ICD-10 approach outlined in this application. While NIHSS is know to have reasonable inter-rater reliability when used in a research setting, it is less clear that it will have similar reliability when applied in claims across all hospitals. It seems plausible that hospitals may have systematic biases in their NIHSS reporting which could substantially influence the validity of hospital level comparisons.

2a2. Reliability Testing

Comments: **The face validity assessment conducted by the developer does not conform to NQF's requirements. I am not convinced that the measure, as specified, indicates quality.

**External validity is somewhat limited analyzing only GWTG hospitals only and by excluding small hospitals. When comparing a registry-only assessment of mortality to the claims-based measure there was very high correlation between these measures. However, as NIHSS is the single biggest predictor of mortality and the same data was used for both measures, it is less persuasive that this establishes concurrent validity as the measures are based, in important part, on the same data.

The sponsors' also cite additional validity arguments — the measure was developed using an approach that is unlikely to lead to bias and the solicitation and incorporation of community feedback is also likely to reduce bias.

The problem with lacking ICD-10 NIHSS severity scores is the question of whether those values can be reliability and accurately measured through a claims-based process that involves all hospitals and not only those hospitals that voluntarily participate in GWTG.

I am uncertain about whether this measure is a measure of quality.

2b2. Validity Testing

Comments: **Exclusions seem fine.

Risk adjustment could be accomplished using either NIHSS or mNIHSS - suggest testing with the mNIHSS, since it is a more reliable measure.

In general, I favor adjustment for SDS factors; in this case, the evidence does not currently support any disparities in care; however, disparities could manifest in future years. I would argue for CT-level SES factors (rather than ZIP-code level).

I am not sure that a difference in IQR of only 1.4 percentage points counts as "substantial".

Multiple imputation methods for estimating NIHSS scores may not be necessary, since the mNIHSS can be estimated from claims data.

**The exclusions have face validity and, given that they are uncommon should not limit external validity.

The risk adjustment variables are all credibly associated with mortality and for the highest leverage variables (age and NIHSS), particularly credible.

Calibration is reasonable, although the model underpredicts at the highest and lowest risks. It is unclear if this has implications for hospital comparisons. The primary concern is that it might slightly bias the measure against hospitals that take care of the highest risk patients.

From a prior theoretical perspective, I would have advocated for including SDS variables in the models. While I would agree that the causal relationship between SDS variables and 30 day

Mortality is unclear, to the extent that associations exists they are unlikely to be related

To hospital quality of care. Moreover there are numerous theoretically credible causal pathways through which these factors may influence mortality (e.g. low income leading to inability to afford medications or receive access to care. Neighbhorhood environment

leading to less ability to obtain assistance or be physically active, etc.)
However, given the empirical fact that SDS measures do not seem to influence the model, it seems reasonable to omit them given the uncertainty of the theoretical argument.
One major problem with the model that is pointed to, however, by considering SDS models comes back to the question of preferences. Why do African Americans have lower mortality? Here, there is a strong prior theoretical argument that this is due to preferences for more intense care amongst African-Americans. If, indeed, that is the explanation for this factor, it speaks to the potential power of preferences in influencing mortality. More importantly, though, it speaks to the potential for the measure to mis-measure quality if preferences are not accounted for and if preferences vary at the hospital level.
The missing data approach is inadequate. First, it is not clear how much missing data will exist when NIHSS is captured through administrative data. Second, multiple imputation is a reasonable approach for correctly estimating model coefficients, but not for estimating hospital level effects. In short, if it is the case (as seems likely) that the degree of missingness varies at the hospital level if there are patient factors that influence mortality and predict missingness then missing data can lead to major biases in hospital estimates of mortality. The fact that missingness correlates with mortality (rather strongly) makes it clear that the data that are missing are missing at random. The implications for this missingness at the hospital level is not clear.

2b3. Exclusions Analysis
2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures
2b5. Identification of Statistically Significant & Meaningful Differences In Performance
2b6. Comparability of Performance Scores When More Than One Set of Specifications
2b7. Missing Data Analysis and Minimizing Bias
Comments: **The results of the measure's testing do not appear to demonstrate sufficient reliability.
**The sample for reliability testing (over 1,000 hospitals) seems appropriately large. One limitation, though, is that the sample may select for artificially high reliability by selecting hospitals already participating in a quality improvement project and excluding hospitals with very low stroke volumes. Even with this potential bias in the measure's favor, reliability was surprisingly poor — in a split sample validation only about half of the variance in each sample was explained by hospital performance in the other half of the sample. While a clear cut standard of acceptable reliability is not obvious, this seems like relatively poor performance for a measure of hospital quality of care.
One possible explanation for this low reliability may be suboptimal reliability in the measurement of NIHSS.

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care. The data are coded by someone other than person obtaining original information.
- Administrative data are routinely collected as part of the billing process. New ICD-10 codes for NIH Stroke Scale scores will be available to hospitals October 2016.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **My concerns about the NIHSS have been previously noted.

**Collection of most data elements is clearly feasible. The questionable element is NIHSS.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details :

- CMS intends to implement this measure in the Hospital Inpatient Quality Reporting (Hospital IQR) Program once the new NIH Stroke Scale ICD-10 codes have been in use for 3 years. This measure requires 3 years of claims data for calculation. Once one of the new measures (either the claims-based or hybrid measure) is implemented it will replace the currently reported Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure.

Potential harms:

- The developer did not identify any unintended consequences related to this measure.

Feedback :

- The measure currently in use, "Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure" was evaluated by NQF in 2012. The lack of inclusion of the NIH Stroke Scale was a major topic during public comment and in committee discussion. The committee could not reach consensus regarding recommending for endorsement a stroke mortality measure that did not include the NIH Stroke Scale value in the risk-adjustment model. The measure was withdrawn from consideration by the measure developer.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **I am skeptical about the usability and use of this measure as specified. Although the measure developer did not identify any unintended consequences, I would argue that using a flawed instrument to measure an outcome that has no demonstrated quality gap would have the unintended consequence of increasing healthcare expenditures (by an estimated $3.50 per case) for no obvious gain.

**While CMS is already publishing a companion measure to this measure, it is not obvious to me how this measure it to be used to further high-quality, efficient healthcare. It does not seem likely that this measure will be used by patients to directly influence quality of care given that stroke is an emergent condition — it seems unlikely that patients will go to a website to determine which is

the best hospital for taking care of stroke. It seems more plausible that it could have effects on influencing hospitals to strive for higher quality care through reputational incentives, but this is also not obvious to me.
I am not aware of any evidence of unintended consequences of the role out of CMS's existing mortality measure, but I am also not aware of any studies that have addressed this question. I do believe that there is a risk for important unintended consequences of this measure, however. Specifically, I am concerned that the measure creates an incentive for hospitals not fully account for patient preferences for mortality over survival with severe disability.

Criterion 5: Related and Competing Measures

Related or competing measures:

- 0467 : Acute Stroke Mortality Rate (IQI 17)
- 2877: Hybrid Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Acute Ischemic Stroke with Risk Adjustment for Stroke Severity
- 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure (not NQF endorsed) - CMS

Harmonization:

- The developers stated that measures' cohorts are harmonized to the extent possible and that the small differences in cohort inclusion and exclusion criteria are appropriate because the measures (0467, 2876, & 2877) assess different stroke outcomes – inpatient mortality vs. 30-day mortality.
- Measure #2877 is also being submitted to NQF for endorsement; this measure uses a combination of claims and electronic health records (EHR) data for risk adjustment. The developer noted that 2877 is otherwise harmonized with this new claims-only measure. It is CMS's intent to implement only one of the new stroke mortality measures (this claims-only measure or the hybrid measure) in any given program.
- In 2012 when the original 30-day stroke mortality measure was submitted and reviewed, as part of their harmonization efforts with other similar mortality measures, the developer re-specified the measure to include all-payer patients ages 18 and over (rather than Medicare FFS patients ages 65+ only). Unlike other 30-day all-cause mortality measures, *this updated version of the stroke mortality measure **does not** include all-payer patients ages 18 and over.*
- It is CMS's intent to replace the current publicly reported measure, Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization, in any given program with this newly developed measure, which includes stroke severity in the risk model.

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following actute ischemic stroke hospitalization with claims-based risk adjustment for stroke severity
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: N/A

Date of Submission: 1/15/2016

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome

☒ Health outcome: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute ischemic stroke hospitalization with claims-based risk adjustment for stroke severity
Health outcome includes patient-reported outcomes (PRO, i.e., HRQoL/functional status, symptom/burden, experience with care, health-related behaviors)

☐ Intermediate clinical outcome: Click here to name the intermediate outcome

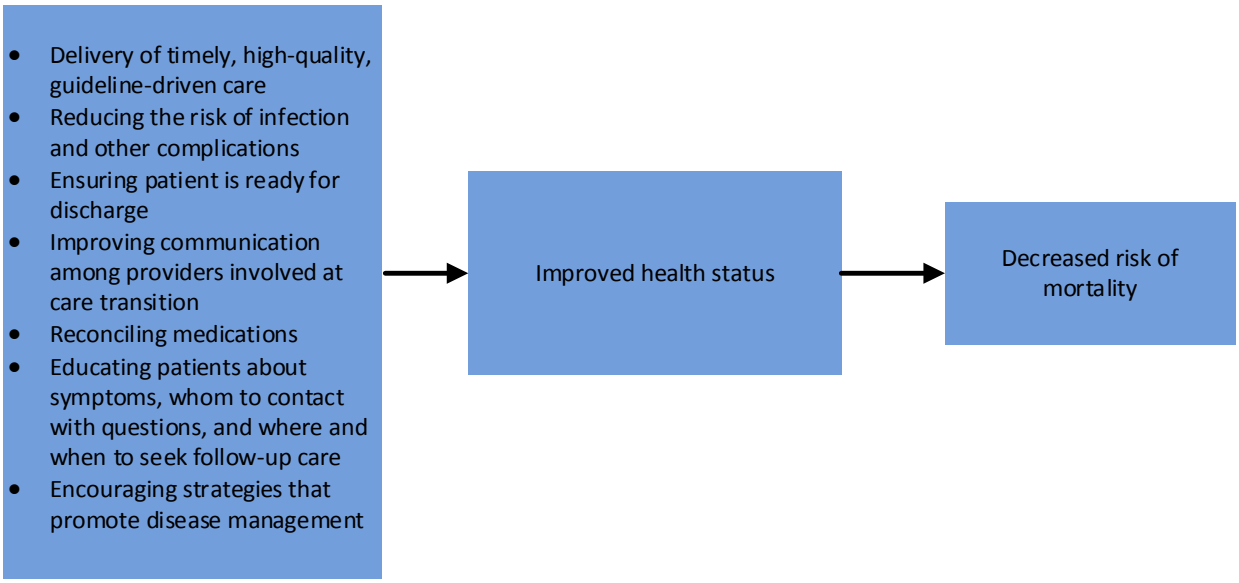
☐ Process: Click here to name the process

☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized mortality rates following hospitalization for stroke. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This mortality measure was developed to identify institutions, whose performance is better or worse than would be expected based on their patient case-mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Stroke is a priority condition for outcomes measure development because it is a leading cause of morbidity and mortality. Stroke affects as many as 795,000 individuals in the United States each year and is the nation's fifth leading cause of death (CDC-NCHS, 2015). It is estimated that stroke costs $34 billion each year in direct and indirect medical costs (Mozaffarian et al., 2015).

Many current hospital processes have been associated with lower stroke mortality rates within 30 days of hospital admission. In particular, post-stroke mortality rates have been shown to be influenced by critical aspects of care at the hospital such as response to complications, speediness of delivery of care, organization of care, and appropriate imaging (Smith et al., 2006; Reeves et al., 2009; Lingsma et al., 2008; Hong et al., 2008; Fonarow et al., 2014). This research demonstrates the relationship between hospital organizational factors and performance on the stroke mortality measure, and supports the ability of hospitals to impact these rates. Stakeholders have previously highlighted the importance of including stroke severity in mortality measures for risk adjustment, as several studies have demonstrated that initial stroke severity is the strongest predictor of mortality in ischemic stroke patients (Smith et al., 2010; Nedeltchev et al., 2010; Fonarow et al., 2012). This update to the current publicly reported measure responds to stakeholder preference to include the National Institutes of Health (NIH) Stroke Scale as an assessment of stroke severity in the risk-adjustment model, thereby accounting for stroke severity at the time of admission to assess the condition of the patient before care has been administered.

Complex and critical aspects of care – such as communication between providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment – all contribute to patient outcomes. For example, hospitals participating in quality improvement registries like Get With The Guidelines (GWTG) had lower in-hospital mortality rates among stroke patients than hospitals not participating in similar programs (Fonarow et al., 2014). Risk-adjusted measures of patient outcomes, specifically mortality, can highlight variations in the provision of care, and thus support improvements by highlighting institutions that provide exceptional care for stroke patients.

The diagram above indicates some of the many care processes that can influence mortality risk. Numerous studies have demonstrated that appropriate (guideline recommended care) and timely treatment for stroke patients can reduce the risk of mortality within 30 days of hospital admission (Hacke et al., 2004; Fang et al., 2008).

References:

CDC, NCHS. Underlying Cause of Death 1999-2013 on CDC WONDER Online Database, released 2015. Data are from the Multiple Cause of Death Files, 1999-2013, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed Feb. 3, 2015.

Fang J, Keenan NL, Ayala C, Dai S, Merritt R, Denny CH. Awareness of stroke warning symptoms—13 states and the District of Columbia, 2005. MMWR. 2008;57(18):481–5.

Fonarow GC, Saver JL, Smith EE, et al. Relationship of national institutes of health stroke scale to 30-day mortality in medicare beneficiaries with acute ischemic stroke. J Am Heart Assoc. Feb 2012;1(1):42-50.

Fonarow GC, Zhao X, Smith EE, et al. Door-to-Needle Times for Tissue Plasminogen Activator Administration and Clinical Outcomes in Acute Ischemic Stroke Before and After a Quality Improvement Initiative. JAMA. 2014;311(16):1632-1640.

Hacke W, Donnan G, Fieschi C, et al. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. Lancet (London, England). Mar 6 2004;363(9411):768-774.

Hong KS, Kang DW, Koo JS, et al. Impact of neurological and medical complications on 3-month outcomes in acute ischaemic stroke. European journal of neurology : the official journal of the European Federation of Neurological Societies. Dec 2008;15(12):1324-1331.

Lingsma HF, Dippel DW, Hoeks SE, et al. Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey. Journal of neurology, neurosurgery, and psychiatry. Aug 2008;79(8):888-894.

Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circulation. 2015 ;e29-322.

Nedeltchev K, Renz N, Karameshev A, et al. Predictors of early mortality after acute ischaemic stroke. Swiss Medical Weekly. 2010;140(17-18):254-259.

Reeves MJ, Smith E, Fonarow G, Hernandez A, Pan W, Schwamm LH. Off-hour admission and in-hospital stroke case fatality in the get with the guidelines-stroke program. Stroke. Feb 2009;40(2):569-576.

Smith EE, Shobha N, Dai D, et al. Risk score for in-hospital ischemic stroke mortality derived andvalidated within the Get With the Guidelines-Stroke Program. Circulation. 2010;122(15):1496-1504.

Smith MA, Liou JI, Frytak JR, Finch MD. 30-day survival and rehospitalization for stroke patients according to physician specialty. Cerebrovascular diseases (Basel, Switzerland). 2006;22(1):21-26.5.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***
☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***
☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***
☐ Other – ***complete section 1a.8***

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***
☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

N/A

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

N/A

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

N/A

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A

1a.7.6. What is the overall quality of evidence <u>across studies</u> in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) <u>across studies</u> in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

_____

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
NQF_2876_Claims-Only_Stroke_Mortality_NQF_Evidence_Attachment_v1.0.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fifth most common cause of death, affecting approximately 795,000 people in the United States annually, and has a mortality rate of 17% [Go et al., 2014; Kochanek et al., 2014]. Stroke is also a leading cause of disability in the United States, which can lead to increased dependency on the health care system and higher subsequent costs associated with this care [Centers for Disease Control and Prevention, 2005]. Mortality following stroke – an important adverse outcome that can be measured reliably and objectively, and that is influenced by the quality of care provided to patients during their initial hospitalization – is an appropriate measure of quality of care [DesHarnais et al., 1988; Weir et al, 2001]. Specifically, post-stroke mortality rates have been shown to be influenced by critical aspects of care such as response to complications, speediness of delivery of care, organization of care, and appropriate imaging [Smith et al., 2006; Reeves et al., 2009; Lingsma et al., 2008; Hong et al., 2008]. This work demonstrates the relationship between hospital organizational factors and performance on the stroke mortality measure and supports the ability of hospitals to impact these rates.

The goal of outcome measurement is to identify institutions whose performance is better or worse than would be expected based on their patient case mix by risk-adjusting for patients' conditions and stroke severity at the time of hospital admission. The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level RSMRs following hospitalization for acute ischemic stroke. Measurement of patient mortality allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures.

Rationale for Development of an Updated Claims-Only Stroke Mortality Measure
Current outcome measures use administrative claims data from the year prior to the index admission in the risk adjustment models. Stakeholders, including the AHA/ASA and other professional organizations, have highlighted the importance of including stroke severity in mortality measures for risk adjustment. Several studies have demonstrated that initial stroke severity is the strongest predictor of mortality in ischemic stroke patients [Smith et al., 2010; Nedeltchev et al., 2010; Fonarow et al., 2012].
This new claims-based stroke mortality measure addresses these stakeholder preferences and improves model performance by updating the current publicly reported claims-based stroke mortality measure to incorporate stroke severity scores into the risk-adjustment model. Advancements in clinical practice to incorporate new clinical assessments in administrative coding systems have made it possible to integrate these data into measures of hospital performance. The NIH Stroke Scale, which was created in 1989 and is widely used in routine stroke care, is collected in the GWTG-Stroke Registry, which has over 1,700 hospitals throughout the U.S. [Fonarow et al., 2014]. The NIH Stroke Scale is a 15-item neurologic examination stroke scale used to provide a quantitative assessment of stroke related neurologic deficit, by evaluating the effect of acute ischemic stroke on the levels of consciousness, language, neglect, visual-field loss, extraocular movement, motor strength, ataxia, dysarthria, and sensory loss. The NIH Stroke Scale is designed to be a simple, valid, and reliable tool that can be administered at the bedside consistently by physicians, nurses, or therapists. The use of the NIH Stroke Scale to assess stroke severity upon acute ischemic stroke patient presentation is recommended in the AHA/ASA Class I guidelines. Furthermore, the NIH Stroke Scale scores will be coded in the ICD-10-CM coding system beginning in October 2016, allowing it to be used in this measure. Inclusion of stroke severity data will not only address stakeholder preferences, but may also improve the discrimination of the risk models.

References:
1. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. Circulation. Jan 21 2014;129(3):e28-e292.

2. Kochanek KD MS, Xu JQ, Arias E. Mortality in the United States, 2013. NCHS data brief, no 178. 2014.
3. Centers for Disease Control and Prevention (CD). Prevalence and most common causes of disability among adults: United States, 2005. MMWR Morb Mortal Wkly Rep. 2009;58:421-426.
4. Casper ML NI, Croft JB, Nilasena DS. Atlas of Stroke Hospitalizations Among Medicare Beneficiaries. 2008.
5. DesHarnais SI, Chesney JD, Wroblewski RT, Fleming ST, McMahon LF, Jr. The Risk-Adjusted Mortality Index. A new measure of hospital performance. Med Care. Dec 1988;26(12):1129-1148.
6. Weir NU, Sandercock PA, Lewis SC, Signorini DF, Warlow CP. Variations between countries in outcome after stroke in the International Stroke Trial (IST). Stroke. Jun 2001;32(6):1370-1377.
7. Smith MA, Liou JI, Frytak JR, Finch MD. 30-day survival and rehospitalization for stroke patients according to physician specialty. Cerebrovascular diseases (Basel, Switzerland). 2006;22(1):21-26.
8. Reeves MJ, Smith E, Fonarow G, Hernandez A, Pan W, Schwamm LH. Off-hour admission and in-hospital stroke case fatality in the get with the guidelines-stroke program. Stroke. Feb 2009;40(2):569-576.
9. Lingsma HF, Dippel DW, Hoeks SE, et al. Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey. Journal of neurology, neurosurgery, and psychiatry. Aug 2008;79(8):888-894.
10. Hong KS, Kang DW, Koo JS, et al. Impact of neurological and medical complications on 3-month outcomes in acute ischaemic stroke. European journal of neurology : the official journal of the European Federation of Neurological Societies. Dec 2008;15(12):1324-1331.
11.Smith EE, Shobha N, Dai D, et al. Risk score for in-hospital ischemic stroke mortality derived and validated within the Get With the Guidelines-Stroke Program. Circulation. 2010;122(15):1496-1504.
12. Nedeltchev K, Renz N, Karameshev A, et al. Predictors of early mortality after acute ischaemic stroke. Swiss Medical Weekly. 2010;140(17-18):254-259.
13. Fonarow GC, Saver JL, Smith EE, et al. Relationship of national institutes of health stroke scale to 30-day mortality in medicare beneficiaries with acute ischemic stroke. J Am Heart Assoc. Feb 2012;1(1):42-50.
14. Adams HP, Jr., Davis PH, Leira EC, et al. Baseline NIH Stroke Scale score strongly predicts outcome after stroke: A report of the Trial of Org 10172 in Acute Stroke Treatment (TOAST).
15. Fonarow GC, Alberts MJ, Broderick JP, et al. Stroke outcomes measures must be appropriately risk adjusted to ensure quality care of patients: a presidential advisory from the American Heart Association/American Stroke Association. Stroke. May 2014;45(5):1589-1601.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*

For model development purposes only, we used two data sources: July 2011-June 2014 Medicare Administrative claims and the 2013 AHA/ASA GWTG-Stroke Registry. Both data sources were linked to create the dataset used for measure development. The registry data served as a surrogate for the NIH Stroke Scale score which, in the future, can be derived from ICD-10 codes in Medicare claims. Our cohort included 188,975 patients at 1,511 hospitals. The mean risk-standardized mortality rate (RSMR) among hospitals was 14.53% and the median hospital RSMR was 14.48%, with a range of 10.75% to 18.98% and an interquartile range was 13.52% to 15.56%.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

Mortality following stroke is an important adverse outcome that can be measured reliably and objectively. The 30-day stroke mortality rate varies by age from 9% in patients 65-74 years of age to 23% in those =85 years of age [Casper et al., 2008]. Post-stroke mortality rates have also been shown to be influenced by critical aspects of care.

Risk-adjusted measures of patient outcomes, including mortality, can highlight variation in the care patients receive across hospitals, and thus support improvements and learning from high quality institutions. The results of CMS's current publicly reported claims-based stroke mortality measure, as reported in the 2014 update to the Hospital Compare, are based on RSMRs calculated for admissions among Medicare FFS patients, age 65 years and older, from July 1, 2010 – June 30, 2013. It includes 520,111 admissions from 4,506 hospitals. The median hospital RSMR was 15.3%, with a range of 8.6% to 23.8%. This variation across hospitals indicates that there is room for improvement in care for stroke patients that could reduce mortality rates.

Additionally, risk-adjusted 30-day mortality rates were shown to decline from 12.1% (95% confidence interval [CI]: 12.0%-12.2%) to 11.6% (95% CI: 11.5%-11.7%) in Medicare FFS from 1999 to 2011 [Krumholz 2014]. This decline suggests that there is opportunity for

further improvement in the 30-day mortality outcome over time.

References:
1. Casper ML NI, Croft JB, Nilasena DS. Atlas of Stroke Hospitalizations Among Medicare Beneficiaries. 2008.
2. Krumholz HM, Normand S-LT, Wang Y. Trends in Hospitalizations and Outcomes for Acute Cardiovascular Disease and Stroke: 1999-2011. Circulation. August 18, 2014 2014.

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Distribution of claims-only stroke RSMRs by proportion of Dual Eligible patients
Dates of data: July 2011 – June 2014
Data source: Medicare FFS claims

Characteristic//Hospitals with a low proportion (0%) Dual Eligible patients//Hospitals with a high proportion (>14.2%) Dual Eligible patients –
Number of Measured Entities (Hospitals)// 551 // 294
Number of Patients// 19,859 in low-proportion hospitals //14,376 in high-proportion hospitals
Maximum// 17.37 // 17.40
90th percentile// 15.06 // 15.12
75th percentile// 14.60 // 14.67
Median (50th percentile)// 14.25 // 14.22
25th percentile// 14.01 // 13.83
10th percentile// 13.60 // 13.39
Minimum // 11.27 // 11.34

Distribution of claims-only stroke RSMRs by proportion of African-American patients
Dates of data: July 2011 – June 2014
Data source: Medicare FFS claims

Characteristic//Hospitals with a low proportion (0%) African-American patients//Hospitals with a high proportion (>17.5%) African-American patients –
Number of Measured Entities (Hospitals)// 492 // 294
Number of Patients// 13,497 in low-proportion hospitals // 17,429 in high-proportion hospitals
Maximum// 17.37 // 17.15
90th percentile// 15.04 // 15.22
75th percentile// 14.60 // 14.65
Median (50th percentile)// 14.27 //14.15
25th percentile// 14.08 //13.71
10th percentile// 13.74 // 13.16
Minimum // 12.15 // 11.27

Distribution of claims-only stroke RSMRs by AHRQ SES Index
Dates of data: July 2011 – June 2014
Data source: Medicare FFS claims and the American Community Survey (2008-2012) data

Characteristic//Hospitals with lowest mean value (=47.68) AHRQ SES Index //Hospitals with highest mean value (>59.38) AHRQ SES Index –
Number of Measured Entities (Hospitals)// 295 // 294
Number of Patients// 11,563 in low-proportion hospitals //22,187 in high-proportion hospitals
Maximum// 17.15 // 16.59
90th percentile// 15.18 // 15.28
75th percentile// 14.66 // 14.67
Median (50th percentile)// 14.27 // 14.17
25th percentile// 13.95 // 13.63
10th percentile// 13.44// 13.14

Minimum // 11.77 // 11.34

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**

Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness

1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**

Stroke is the fifth most common cause of death in the United States each year with approximately 795,000 people having a stroke, and has a mortality rate of 17% [Go et al., 2014]. Stroke is one of CMS's top 10 costliest conditions [Go et al., 2014; Total Expenses, 2012]. The direct medical costs of stroke in 2010 (which includes hospital outpatient visits, hospital inpatient stays, emergency department visits, prescribed medications, and home health care) was estimated to be $20.6 billion [Total Expenses, 2012]. This is also a condition for which there is room for performance improvement in mortality outcomes as discussed in 1b.3.

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. Circulation. Jan 21 2014;129(3):e28-e292.
2. Total Expenses and Percent Distribution for Selected Conditions by Source of Payment: United States, 2012. Medical Expenditure Panel Survey Household Component Data. Generated interactively. 2012.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

N/A. This measure is not a PRO-PM.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*
Neurology, Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
Care Coordination, Safety, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:** Testing_Form_Claims-Only_Stroke_Mortality_01-05-16_Finaldocx_sr.docx

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** NQF_2876_Claims-Only_Stroke_Mortality_S2b_Mortality_Data_Dictionary_v1.0-635884757617681755.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.
N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The outcome for this measure is 30-day, all-cause mortality. We define mortality as death from any cause within 30 days of the index admission for patients with a principal discharge diagnosis of acute ischemic stroke.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator time window: We define mortality as death from any cause within 30 days of the index admission.
Denominator time window: This measure will use index admissions during a three-year period for the denominator.

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The measure outcome is death from any cause within 30 days of the index admission date. As currently specified, we identify deaths for FFS Medicare patients, age 65 years and older, in the Medicare Enrollment Database (EDB).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The cohort includes inpatient admissions to all non-federal, short-term, acute care hospitals for Medicare FFS patients age 65 years and older with a principal discharge diagnosis of acute ischemic stroke.

Additional details are provided in S.9 Denominator Details.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The denominator includes all Medicare FFS beneficiaries, age 65 and over, with a principal discharge diagnosis of acute ischemic stroke. To be included in the measure cohort used in public reporting, patients must meet the following additional inclusion criteria:

1. Enrolled in Medicare fee-for-service (FFS) during the index admission;
2. Not transferred from another acute care facility; and
3. Enrolled in Part A and Part B Medicare for the 12 months prior to the date of index admission.

ICD-9-CM codes that define the patient cohort:
433.01 Occlusion and stenosis of basilar artery with cerebral infarction
433.11 Occlusion and stenosis of carotid artery with cerebral infarction
433.21 Occlusion and stenosis of vertebral artery with cerebral infarction
433.31 Occlusion and stenosis of multiple and bilateral precerebral arteries with cerebral infarction
433.81 Occlusion and stenosis of other specified precerebral artery with cerebral infarction
433.91 Occlusion and stenosis of unspecified precerebral artery with cerebral infarction
434.01 Cerebral thrombosis with cerebral infarction
434.11 Cerebral embolism with cerebral infarction
434.91 Cerebral artery occlusion, unspecified with cerebral infarction
436 Acute, but ill-defined, cerebrovascular disease

ICD-10 codes that define the patient cohort:
I63.22 Cerebral infarction due to unspecified occlusion or stenosis of basilar arteries
I63.139 Cerebral infarction due to embolism of unspecified carotid artery
I63.239 Cerebral infarction due to unspecified occlusion or stenosis of unspecified carotid arteries
I63.019 Cerebral infarction due to thrombosis of unspecified vertebral artery
I63.119 Cerebral infarction due to embolism of unspecified vertebral artery
I63.219 Cerebral infarction due to unspecified occlusion or stenosis of unspecified vertebral arteries
I63.59 Cerebral infarction due to unspecified occlusion or stenosis of other cerebral artery
I63.20 Cerebral infarction due to unspecified occlusion or stenosis of unspecified precerebral arteries
I63.30 Cerebral infarction due to thrombosis of unspecified cerebral artery
I63.40 Cerebral infarction due to embolism of unspecified cerebral artery
I63.50 Cerebral infarction due to unspecified occlusion or stenosis of unspecified cerebral artery
I67.8 Other specified cerebrovascular diseases
I67.89 Other cerebrovascular diseases

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes admissions for patients:
1. With inconsistent or unknown vital status or other unreliable data;
2. Enrolled in the Medicare hospice program at any time in the 12 months prior to the index admission, including the first day of the index admission; and
3. Discharged against medical advice (AMA).
For patients with more than one admission for stroke in a given year, only one index admission for that condition is randomly selected for inclusion in the cohort.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

1. Inconsistent vital status or unreliable data: We do not include stays for patients where the age is greater than 115, where the gender is neither male nor female, where the admission date is after the date of death in the Medicare Enrollment Database, or where the date of death occurs before the date of discharge but the patient was discharged alive.

2. Hospice enrollment in the 12 months prior to or on the index admission is identified using hospice data and the Inpatient Standard Analytic File (SAF). These patients are likely continuing to seek comfort measures only; thus, mortality is not necessarily an adverse outcome or signal of poor quality care for these patients.

3. Discharges against medical advice (AMA) are identified using the discharge disposition indicator. After all exclusions are applied, the measure randomly selects one index admission per patient per year for inclusion in the cohort so that each episode of care is mutually independent with the same probability of the outcome. For each patient, the probability of death increases with each subsequent admission, and therefore, the episodes of care are not mutually independent. Similarly, for the three year combined data, when index admissions occur during the transition between measure reporting periods (June and July of each year) and both are randomly selected for inclusion in the measure, the measure includes only the June admission. The July admissions are excluded to avoid assigning a single death to two admissions.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*
Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).

The measure employs a hierarchical logistic regression model to create a hospital-level 30-day RSMR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, it models the log-odds of hospital mortality within 30 days of discharge using age, selected clinical covariates, and a hospital-specific intercept. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of mortality at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Candidate and Final Risk-adjustment Variables:
Candidate variables are patient-level risk-adjustors that are expected to be predictive of mortality, based on empirical analysis, prior literature, and clinical judgment, including age, sex, and indicators of comorbidity and disease severity. Only claims variables in the current publically reported claims-based stroke mortality measure are considered as candidate variables. For each patient, covariates are obtained from claims records extending 12 months prior to and including the index admission. For this measure, these risk-adjustment variables are identified using both inpatient and outpatient Medicare FFS claims data. The model adjusts for case-mix differences based on the clinical status of patients at the time of admission. We use condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care and that are only recorded in the index admission. This list includes a variable indicating whether a patient is transferred into an inpatient admission from the emergency department (ED). The ED transfer indicator variable is also a risk-adjustment variable derived from claims that is used in the current publically reported claims-based stroke mortality measure. It was added to the risk-adjustment model for that measure in 2013 because these patients are believed to have a higher risk of mortality. Therefore, we include this variable in the list of candidate variables for the hybrid stroke mortality measure as well

We also include the NIH Stroke Scale in our list of candidate variables. We used NIH Stroke Scale score values collected in the American Heart Association/American Stroke Association (AHA/ASA) Get With The Guidelines (GWTG)-Stroke Registry in order to include a measure of stroke severity in the list of candidate variables.

The final set of risk adjustment variables is:

Demographics
Age-65 (continuous, per 5 years)

Additional Risk Factors
Transfer from another ED
NIH Stroke Scale score (continuous, per 5 units)

Comorbidities
Congestive heart failure (CC 80)
Congenital cardiac/circulatory defects (CC 87-88)
Specified heart arrhythmias (CC 92)
Cerebral atherosclerosis and aneurysm (CC 98)
Metastatic cancer and acute leukemia and other major cancers (CC 7-8)
Protein-calorie malnutrition (CC 21)
Other significant endocrine and metabolic disorders (CC 22-24)
Other gastrointestinal disorders (CC 36)
Disorders of the vertebrae and spinal discs (CC 39)
Osteoarthritis of hip or knee (CC 40)
Other musculoskeletal and connective tissue disorders (CC 43)
Iron deficiency and other/unspecified anemia and blood disease (CC 47)
Dementia or other specified brain disorders (CC 49-50)
Multiple sclerosis (CC 72, 76)
Seizure disorders and convulsions (CC 74)
Pneumonia (CC 111-113)
Renal failure (CC 131)

References:
Krumholz HM, Brindis RG, Brush JE, et al. 2006. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation 113: 456-462.

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)

S.16. Type of score:
Rate/proportion
If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score,*

a lower score, a score falling within a defined interval, or a passing score)
Better quality = Lower score

S.18. **Calculation Algorithm/Measure Logic** *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*
The measure estimates hospital-level, 30-day, all-cause RSMRs following hospitalization for stroke using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of mortality within 30 days of index admission using age, selected clinical covariates, and a hospital-specific intercept. At the hospital level, it models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a mortality at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

The RSMR is calculated as the ratio of the number of "predicted" to the number of "expected" deaths at a given hospital, multiplied by the national observed mortality rate. For each hospital, the numerator of the ratio is the number of deaths within 30 days predicted on the basis of the hospital's performance with its observed case mix, and the denominator is the number of deaths expected based on the nation's performance with that hospital's case mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows for a comparison of a particular hospital's performance given its case mix to an average hospital's performance with the same case mix. Thus, a lower ratio indicates lower-than-expected mortality rates or better quality, and a higher ratio indicates higher-than-expected mortality rates or worse quality.

The "predicted" number of deaths (the numerator) is calculated by using the coefficients estimated by regressing the risk factors and the hospital-specific intercept on the risk of mortality. The estimated hospital-specific intercept is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. The results are transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of deaths (the denominator) is obtained in the same manner, but a common intercept using all hospitals in our sample is added in place of the hospital-specific intercept. The results are transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimate the model coefficients using the years of data in that period.

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed mortality rate. The hierarchical logistic regression models are described fully in the original methodology report (Grosso et al., 2011).

References:
Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A. This measure is not based on a sample or survey.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A. This measure is not based on a sample or survey.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Missing values are rare among variables used from claims data in this measure. However, for measure development, the NIH Stroke Scale score was pulled from the AHA/ASA GWTG-Stroke Registry as a surrogate for future claims-based data. Collection of the NIH Stroke Scale score is now recommended in the AHA/ASA guidelines as Class I for care of patients admitted with acute ischemic stroke. Analysis of data from the AHA/ASA GWTG-Stroke Registry indicate that hospitals are collecting these data on an increasing

proportion of patients admitted for acute ischemic stroke. Among the measure cohort, NIH Stroke Scale scores were available in 82.39% of patients with an admission for ischemic stroke from July 1, 2013 and June 30, 2014. This proportion increased from 70.78% of admissions occurring between July 1, 2011 and June 30, 2012. It is CMS's intention to use claims data for NIH Stroke Scale scores when this measures is implemented. New ICD-10 codes for NIH Stroke Scale scores will be available for use in October 2016. Reliability of the NIH Stroke Scale data will be reassessed prior to implementation and after hospitals begin using these codes in Medicare claims.

To account for missing values of the NIH Stroke Scale score during measure development, we used multiple imputation to generate a range of plausible values for all missing data and estimate values for missing data. In multiple imputation, missing variable values are predicted using other patient variables available. The predicted values are substituted for the missing values, which results in a full data set (the imputed data set) without any missing variables. By repeating this process multiple times, we get multiple imputed data sets. We then conduct analyses on and obtain results for each imputed data set. The results based on multiple data sets are combined to produce the overall final results. Because we do not rely on one particular plausible version of the value, we have multiple versions of the plausible values. In general, imputed values are not intended to be "guesses" of what any particular missing value might be; instead, multiple imputation is used to preserve the important characteristics of the underlying data set and the inherent relationships among the variables in the data set. The multiple imputation represents a random sample of the missing values according to the association of the non-missing values of all the variables considered. The resulting inferences of multiple imputation are statistically valid and reflective of the uncertainty due to missing values [He & Belin, 2014; Carpenter & Kenward; Rubin, 1987].

Five copies of imputation datasets were produced for the analyses, and then the results based on these data separately were aggregated according to the standard statistical methods for presentation and for the measure score calculation.

For measure implementation, we anticipate missing values to be rare since the NIH Stroke Scale score will be derived from Medicare claims data. CMS has not made any decisions about how missing data will be handled when the measure is implemented.

References:

He R, Belin T. Multiple imputation for high-dimensional mixed incomplete continuous and binary data. Stat. Med. 2014;33:2251–2262.
Carpenter J, Kenward M. Wiley: Multiple Imputation and its Application - [Internet]. [cited 2015 May 18];Available from: http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470740523.html
Rubin DB. Frontmatter [Internet]. In: Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, Inc.; 1987 [cited 2015 May 15]. p. i–xxix.Available from: http://onlinelibrary.wiley.com/doi/10.1002/9780470316696.fmatter/summary

S.23. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Registry, Other

S.24. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
For measure implementation the data sources will be:
1. Medicare Part A inpatient and Part B outpatient claims: This data source contains claims data for fee-for service inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, as well as inpatient and outpatient physician claims for the 12 months prior to an index admission.
2. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission, as well as vital status. These data have previously been shown to accurately reflect patient vital status (Fleming et al., 1992).
3. For measure development purposes only, we linked the data sources above with data from the AHA/ASA GWTG-Stroke Registry. The registry data were used to obtain the National Institutes of Health (NIH) Stroke Scale scores and clinical risk variables. When this measure is implemented NIH Stroke Scale scores will be derived from ICD-10 codes in Medicare claims.

Reference:

Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs hospitals. Medical Care. 1992; 30(5): 377-91. Data sources for the all-payer

update

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF_2876_Claims-Only_Stroke_Mortality_NQF_Testing_Attachment_v1.1.docx

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute ischemic stroke hospitalization with claims-based risk adjustment for stroke severity

Date of Submission: 1/14/2016

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed

separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: **(*must be consistent with data sources entered in S.23*)**	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Census Data/American Community Survey	☒ other: Census Data/American Community Survey

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

The dataset used for testing included Medicare Parts A and B claims, the Medicare Enrollment Database (EDB), and data from the American Heart Association/American Stroke Association (AHA/ASA)'s Get With The Guidelines (GWTG)-Stroke Registry. The registry data were used as a surrogate for data that will eventually come from claims once ICD-10 codes for stroke severity are available and consistently used by hospitals. The registry data were also used for measure validation. Additionally, census as well as claims data were used to assess socioeconomic factors and race (dual eligible and African American race variables obtained through enrollment data; Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score obtained through census data). The dataset used varies by testing type; see Section 1.7 for details.

1.3. What are the dates of the data used in testing?

The dates used vary by testing type; see Section 1.7 for details.

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.26*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other:

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

For this measure, hospitals are the measured entities. The testing dataset included a subset of 1,520 non-federal, acute inpatient US hospitals (including territories) that participate in the American Heart Association/American Stroke Association (AHA/ASA)'s Get With The Guidelines (GWTG)-Stroke Registry. The number of measured entities varies slightly by the type of testing performed; see Section 1.7 for details.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

The number of admissions/patients varies by testing type; see Section 1.7 for details.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

The datasets, dates, number of measured hospitals, and number of admissions used in each type of testing are as follows:

For reliability testing

The reliability of the model was tested by randomly selecting 50% of **Dataset 1** (development dataset) and developing a risk-adjusted model for this group. We then developed a second model for the remaining 50% of patients (validation cohort) and compared the two. Thus, for reliability testing, we randomly split **Dataset 1** into two samples.

For measure development purposes only, we used two linked data sources to create **Dataset 1**: Medicare Administrative claims and the AHA/ASA GWTG-Stroke Registry. Registry data were used to obtain the National Institutes of Health (NIH) Stroke Scale scores that could in the future be obtained from ICD-10 claims data.

Dataset 1 (development dataset): Merged Medicare Part A Inpatient and Outpatient and Part B Outpatient claims with GWTG-Stroke Registry stroke severity data
Dates of Data: July 1, 2011 – June 30, 2014
Number of Admissions: 188,975
Patient Descriptive Characteristics: average age=79.47, %male=43.39
Number of Measured Hospitals: 1,511

First half of split sample (development sample):
- Number of Admissions: 94,466
- Number of Measured Hospitals: 1,473

Second half of split sample (validation sample):
- Number of Admissions: 94,509
- Number of Measured Hospitals: 1,462

For validity testing (Section 2b2)
Dataset 1

For testing of measure exclusions (Section 2b3)
Dataset 1 (prior to exclusions being applied)
Number of Eligible Admissions: 217,723
Number of Eligible Measured Entities: 1,520

For testing of measure risk adjustment (Section 2b4)
Dataset 1 development and validation samples

For testing to identify meaningful differences in performance (Section 2b5)
Dataset 1

For testing of socioeconomic status (SES) factors and race in risk models (Section 2b4.4b)
Dataset 1 and **Dataset 2**: The American Community Survey (2008-2012)

We examined disparities in performance according to the proportion of patients in each hospital who were of African-American race and the proportion who were dual eligible for both Medicare and Medicaid insurances. We also used the AHRQ SES index score derived from the American Community Survey (2008-2012) (**Dataset 2**) to study the association between performance measures and socioeconomic status.

Data Elements
- African-American race and dual eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data are obtained from CMS enrollment data (**Dataset 1**)

• Validated AHRQ SES index score is a composite of 7 different variables found in the census data (**Dataset 2**, the American Community Survey [2008-2012])

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Sociodemographic status incorporates socioeconomic variables as well as race into a more concise term. However, given the fact that socioeconomic risk factors are distinct from race and should be interpreted differently, we have decided to refer to "socioeconomic status" and "race" separately throughout this form.

We selected socioeconomic status (SES) and race variables to analyze after reviewing the literature and examining available national data sources. There is a large body of literature linking various SES factors and African-American race to worse health status and higher mortality over a lifetime (Adler and Newman, 2002; Mackenbach et al., 2000; Tonne et al., 2005; van Oeffelen et al., 2012). Income, education, and occupational level are the most commonly examined variables. Studies examining stroke mortality have suggested an association with race and SES factors (Khan et al., 2011; Pedigo et al., 2011; Glymour et al., 2009; Clark et al., 2011; Boan et al., 2014, Howard et al., 2011); however, the literature contains few studies directly examining how different SES factors or race might influence the likelihood of older, insured, Medicare patients dying within 30 days of an admission for a stroke. The causal pathways for SES and race variable selection are described below in Section 2b4.3.

The SES and race variables used for analysis were:

- Dual eligible status (**Dataset 1**)
- African-American race (**Dataset 1**)
- AHRQ-validated SES index score using 5-digit zip code data (percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th-grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room) (**Dataset 2**)

References:

Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. *Health affairs (Project Hope).* 2002; 21(2):60-76.

Boan AD, Feng WW, Ovbiagele B, et al. Persistent racial disparity in stroke hospitalization and economic impact in young adults in the buckle of stroke belt. Stroke; a journal of cerebral circulation. Jul 2014;45(7):1932-1938.

Clark CJ, Guo H, Lunos S, et al. Neighborhood cohesion is associated with reduced risk of stroke mortality. Stroke; a journal of cerebral circulation. May 2011;42(5):1212-1217.

Glymour MM, Kosheleva A, Boden-Albala B. Birth and adult residence in the Stroke Belt independently predict stroke mortality. Neurology. Dec 1 2009;73(22):1858-1865.

Howard VJ, Kleindorfer DO, Judd SE, et al. Disparities in stroke incidence contributing to disparities in stroke mortality. Ann Neurol 2011;69:619–627.

Khan JA, Casper M, Asimos AW, et al. Geographic and sociodemographic disparities in drive times to Joint Commission-certified primary stroke centers in North Carolina, South Carolina, and Georgia. Preventing chronic disease. Jul 2011;8(4):A79.

Mackenbach JP, Cavelaars AE, Kunst AE, Groenhof F. Socioeconomic inequalities in cardiovascular disease mortality; an international study. *European heart journal.* 2000; 21(14):1141-1151.

Pedigo A, Seaver W, Odoi A. Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: fuzzy cluster and discriminant analyses approaches. PloS one. 2011;6(7):e22693.

Tonne C, Schwartz J, Mittleman M, Melly S, Suh H, Goldberg R. Long-term survival after acute myocardial infarction is lower in more deprived neighborhoods. *Circulation.* Jun 14 2005; 111(23):3063-3070.

van Oeffelen AA, Agyemang C, Bots ML, et al. The relation between socioeconomic status and short-term mortality after acute myocardial infarction persists in the elderly: results from a nationwide study. *European journal of epidemiology.* Aug 2012; 27(8):605-613.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Data Element Reliability

In constructing the measure, we aim to utilize only those data elements from the claims that have both face validity and reliability. We generally avoid the use of fields that are thought to be coded inconsistently across hospitals or providers. Specifically, we use fields that are consequential for payment and which are audited. We identify such variables through empiric analyses and our understanding of CMS auditing and billing policies and seek to avoid variables which do not meet this standard. For example, "discharge disposition" is a variable in Medicare claims data that is not thought to be a reliable variable for identifying a transfer between two acute care facilities. Thus, we derive a variable using admission and discharge dates as a surrogate for "discharge disposition" to identify hospital admissions involving transfers. This allows us to identify these admissions using variables in the claims data which have greater reliability than the "discharge disposition" variable. In addition, CMS has in place several hospital auditing programs used to assess overall claims code accuracy, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in our measures, including diagnosis and procedure codes and other elements that are consequential to payment.

In addition to comorbidity data from claims, this measure includes the NIH Stroke Scale score to adjust for stroke severity, which will be included in ICD-10 CM codes beginning in October 2016. For development, we used NIH Stroke Scale data from the American Heart Association/American Stroke Association (AHA/ASA)'s Get With The Guidelines (GWTG)-Stroke Registry. However, it is CMS's intention to use claims data, and not registry data, for NIH Stroke Scale scores when this measures is implemented. New ICD-10 codes for NIH Stroke Scale scores will be available for use in October 2016. Given clinical guidelines and extensive prior testing and use of this score we expect reliable coding of stroke severity, but CMS will monitor its use as a part of this measure implementation.

Finally, we assess the reliability of the data elements by comparing model variable frequencies and odds ratios from logistic regression models in the development and validation samples (July 1, 2011-June 30, 2014, **Dataset 1**).

Measure Score Reliability

The reliability of a measurement is the degree to which repeated measurements of the same entity agree with each other. For measures of hospital performance, the measured entity is the hospital, and reliability is the extent to which repeated measurements of the same hospital give similar results. In line with this thinking, our approach to assessing reliability is to consider the extent to which assessments of a hospital using different but randomly selected subsets of

patients produces similar measures of hospital performance. That is, we take a "test-retest" approach in which hospital performance is measured once using a random subset of patients, then measured again using a second random subset exclusive of the first, and finally comparing the agreement between the two resulting performance measures across hospitals (Rousson et al., 2002).

For test-retest reliability, we combined index admissions from successive measurement periods into one dataset, randomly sampled half of patients within each hospital, calculated the measure for each hospital, and repeated the calculation using the second half. Thus, each hospital is measured twice, but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of these two subsets agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement we calculated the intra-class correlation coefficient (ICC), and assessed the values according to conventional standards (Landis and Koch, 1977; Shrout and Fleiss, 1979). Specifically, we used **Dataset 1** split samples and calculated the RSMR for each hospital for each sample. The agreement of the two RSMRs was quantified for hospitals in each sample using the intra-class correlation as defined by ICC (2,1) by Shrout and Fleiss (1979).

Using two independent samples provides a stringent estimate of the measure's reliability, compared with using two random but potentially overlapping samples which would exaggerate the agreement. Moreover, because our final measure is derived using hierarchical logistic regression, and a known property of hierarchical logistic regression models is that smaller volume hospitals contribute less 'signal,' a split sample using a single measurement period would introduce extra noise. This leads to an underestimate in the actual test-retest reliability that would be achieved if the measure were reported using the full measurement period, as evidenced by the Spearman Brown prophecy formula (Spearman, 1910; Brown, 1910). We use this to estimate the reliability of the measure if the whole cohort were used, based on an estimate from half the cohort.

References:

Brown, W. (1910). Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3, 296–322.

Landis J, Koch G, The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.

Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test–retest reliability of continuous measurements. Statistics in Medicine 2002;21:3431-3446.

Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979;86:420-428.

Spearman, Charles, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271–295.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data element reliability results (**Dataset 1**)

The frequency of model variables remained relatively constant between 2011 and 2014, with no model variables increasing or decreasing by more than 2%.

Analysis of data from the Get With The Guidelines (GWTG)-Stroke Registry indicate that hospitals are collecting these data with increasing frequency for patients admitted for acute ischemic stroke. Among the measure cohort, NIH Stroke Scale scores were available in 82.39% of patients with an admission for ischemic stroke from July 1, 2013 and June 30, 2014. This proportion increased from 70.78% of admissions occurring between July 1, 2011 and June 30, 2012.

Examination of the odds ratios for each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across three years (see attached "Hospital-Level Measures of 30-Day Mortality Following Acute Ischemic Stroke Hospitalization that Incorporate Risk Adjustment for Stroke Severity Technical Report").

For the model variable frequencies and risk variable odds ratios, see field S.2b (Data Dictionary or Code Table).

Measure Score Reliability Results (**Dataset 1**)

There were 188,975 admissions in the measure cohort, with 94,466 in one randomly selected sample (development sample) and 94,509 in the other sample (validation sample). The agreement between the two RSMRs for each hospital was 0.556, which according to the conventional interpretation is "moderate" (Landis & Koch, 1977).

Note that this analysis was limited to hospitals with 12 or more cases in each split sample. The intra-class correlation coefficient is based on a split sample of three years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data, whereas the measure is intended to be reported with the full three years of data.

Reference:

Landis J, Koch G. The measurement of observer agreement for categorical data, Biometrics 1977;33:159-174.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The stability over time of the odds ratios and model variable frequencies suggests that the underlying data elements are reliable. The ICC demonstrates moderate agreement across samples using a conservative approach to assessment for the measure score.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

- ☒ **Empirical validity testing**
- ☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

During measure development, we validated the administrative risk model with a risk model that uses data from the GWTG-Stroke Registry derived from patients' medical records. Measure validity is also demonstrated through use of established measure development guidelines.

Validation against Registry Data (**Dataset 1**):

We validated the stroke mortality administrative model, which uses claims-based data, against a model which uses GWTG- Stroke Registry data derived from patients' medical records. For the derivation of the registry model, we linked cases from the registry to the corresponding administrative data from Medicare claims and mortality data from the Medicare enrollment database (**Dataset 1**). The GWTG-Stroke Registry draws data from medical records and has been shown to be reliable through the studies comparing registry data to chart abstraction (Xian et al., 2012). The measure cohort used for the registry model used inclusion/exclusion criteria and risk-adjustment strategy that was consistent with the claims-based administrative model, but the registry model used chart-based risk adjusters, such as blood pressure, which were not available in the claims data. Because only patients aged 65 years and older were included, and some data were excluded based on linkage and other factors, a total of 188,975 stroke hospitalizations were used in the analysis. We compared the hospital performance results of the two models.

Validity Indicated by Established Measure Development Guidelines:

We developed this measure in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcomes measurement set forth in NQF guidance for outcomes measures (National Quality Forum, 2010), CMS Measure Management System

(MMS) guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz, Brindis, et al., 2006).

Validity as Assessed by External Groups:

Throughout measure development, we obtained expert and stakeholder input via regular discussions with an advisory working group and a 30-day public comment period in order to increase transparency and to gain broader input into the measure.

The working group was assembled, and regular meetings were held throughout the development phase. The working group was tailored for development of this measure and consisted of clinicians (neurologists and cardiologists) and other professionals with expertise in biostatistics, measure methodology, and quality improvement. The working group meetings addressed key issues related to measure development, including the deliberation and finalization of key decisions (e.g., defining the measure cohort and outcome) to ensure the measure is meaningful, useful, and well-designed. The working group provided a forum for focused expert review and discussion of technical issues during measure development.

Following completion of the preliminary model, we solicited public comment on the measure through the CMS site link: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html. The public comments were then posted publicly for 30 days. The resulting input was taken into consideration during the final stages of measure development and contributed to minor modifications to the measure.

Citations:

Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. *PLoS One* 2011;6(4):e17401.

Keenan PS, Normand SL, Lin Z, et al. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. *Circulation* 2008;1(1):29-37.

Krumholz HM, Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation.* January 24, 2006 2006;113(3):456-462.

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation* 2006;113(13):1683-92.

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. *Circulation* 2006;113:1693-1701.

National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx. Accessed August 19, 2010.

Shahian DM, He X, O'Brien S, et al. Development of a Clinical Registry-Based 30-Day Readmission Measure for Coronary Artery Bypass Grafting Surgery. Circulation 2014; DOI: 0.1161/CIRCULATIONAHA.113.007541. Published online before print June 10, 2014

Suter L, Wang C, Araas M, et al. Hospital-Level 30-Day All-Cause Unplanned Readmission Following Coronary Artery Bypass Graft Surgery (CABG): Updated Measure Methodology Report. 2014; http://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890352615&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmsn_CABG_MeasMethd_Rpt_060314.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed November 4, 2015.

Xian Y, Fonarow GC, Reeves MJ, et al. Data quality in the American Heart Association Get With The Guidelines-Stroke (GWTG-Stroke): Results from a National Data Validation Audit. *American Heart Journal.* 2012;163(3):392-398.e391. http://www.ahjonline.com/article/S0002-8703%2811%2900894-5/abstract

ICD-9 to ICD-10 Conversion

Statement of Intent

[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.

[ ] The intent of the measure has changed.

Process of Conversion

ICD-10 codes were identified using 2015 GEM mapping software. We enlisted the help of clinicians with expertise in relevant areas to select and evaluate which ICD-10 codes map to the ICD-9 codes currently in use for this measure. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Registry Validation

The performance of the administrative and GWTG-Stroke Registry models is similar (**Dataset 1**). The areas under the receiver operating characteristic (ROC) curve are 0.8120 and 0.7939, respectively, for the two models.

We estimated hospital-level RSMRs using the corresponding hierarchical logistic regression administrative and registry models for the linked patient sample. We then examined the linear relationship between the two sets of estimates using regression techniques and weighting by the total number of cases in each hospital. The Pearson correlation coefficient, weighted by hospital volume, of the standardized rates from the administrative and registry models is 0.95647.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Registry Validation (**Dataset 1**)

The results between the administrative and registry models proved to be similar in each of the model testing that was performed. The ROC results were nearly identical and in line with other mortality models. The Pearson correlation coefficient, weighted by hospital volume, shows the resulting measure from the administrative claims model is as good as that from the registry model.

Validity as Assessed by External Groups

The face validity testing results demonstrated working group and public comment agreement with overall face validity of the measure as specified.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Dataset 1**). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

In **Dataset 1** (prior to exclusions being applied):

Exclusion	N	%	Distribution across hospitals (N=1,511): Min, 25th, 50th, 75th percentile, max
1. Inconsistent or unknown vital status or other unreliable data	0	0.0%	(0, 0, 0, 0, 0)
2. Enrolled in the Medicare hospice program at any time in the 12 months prior to the index admission, including the first day of the index admission	1,527	0.78%	(0, 0, 0, 1, 15)
3. Discharged against medical advice (AMA)	439	0.22%	(0, 0, 0, 0, 6)

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusions 1 and 2 are necessary for valid calculation of the measure. **Exclusion 1** (patients with inconsistent or unknown vital status or other unreliable demographic [age and gender] data) accounts for 0.0% of all index admissions excluded from the initial index cohort. We do not include stays for patients where the age is greater than 115, where the gender is neither male nor female, where the admission date is after the date of death in the Medicare Enrollment Database, or where the date of death occurs before the date of discharge but the patient was discharged alive. **Exclusion 2** (patients enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including the first day of the index admission) accounts for 0.78% of all index admissions excluded from the initial index cohort. These patients are likely continuing to seek comfort measures only; mortality is not necessarily an adverse outcome or signal of poor quality care.

Exclusion 3 (patients who are discharged AMA) accounts for 0.22% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to deliver full care and prepare the patient for discharge.

After all exclusions are applied, the measure randomly selects one index admission per patient per year for inclusion in the cohort so that each episode of care is mutually independent with the same probability of the outcome. For each patient, the probability of death increases with each subsequent admission, and therefore, the episodes of care are not mutually independent. Similarly, for the three year combined data, when index admissions occur during the transition between measure reporting periods (June and July of each year) and both are randomly selected for inclusion in the measure, the measure includes only the first admission. The subsequent admission is excluded to avoid assigning a single death to two admissions.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with** 20 **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Our approach to risk adjustment was tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).

The measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital-level 30-day RSMR. This approach to modeling appropriately accounts for the structure of the data (patients clustered within hospitals), the underlying risk due to patients' comorbidities, and sample size at a given hospital when estimating hospital mortality rates. In brief, the approach simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals (Normand and Shahian et al., 2007). At the patient level, each model adjusts the log-odds of mortality within 30-days of admission for age, selected clinical covariates and a hospital-specific intercept. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept, or hospital-specific effect, represents the hospital contribution to the risk of mortality, after accounting for patient risk and sample size, and can be inferred as a measure of quality. The hospital-specific intercepts are given a distribution in order to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Clinical Factors

We sought to develop a model that included key variables that were clinically relevant and based on strong relationships with the mortality outcome. We also sought to develop a model that was parsimonious, using a grouper that is in the public domain for the 15,000+ ICD-9-CM codes. The candidate variables for the model were derived from: the index admission, with comorbidities identified from the index admission secondary diagnoses (excluding potential complications); 12-month pre-index inpatient data (for any condition); outpatient hospital data; Part B physician data; and NIH Stroke Scale from the GWTG-Stroke Registry. The registry NIH Stroke Scale data were used as a surrogate for data that will eventually come from claims once ICD-10 codes for stroke severity are available and consistently used by hospitals. We developed candidate variables for the model from the claims codes.

For administrative claims model development, we generally begin with the 189 diagnostic groups included in the Hierarchical Condition Category (HCC) clinical classification system (Pope et al., 2000). The HCC clinical classification system was developed for CMS in preparation for all-encounter risk adjustment for Medicare Advantage (managed care) plans and represented a refinement of an earlier risk-adjustment method based solely on principal inpatient diagnosis. The HCC model makes use of all physician and hospital encounter diagnoses and was designed to predict a beneficiary's expenditures based on the total clinical profile represented by all of his/her assigned HCCs. Under the HCC algorithm, the 15,000+ ICD-9-CM diagnosis codes are first assigned to one of 804 mutually exclusive groupings ("DxGroups") and then subsequently aggregated into 189 condition categories (CCs). During development, we used the April 2008 version of the ICD-9-CM to CC assignment map, which is maintained annually by CMS and posted at www.qualitynet.org. We do not use the hierarchy and therefore refer to the CCs rather than HCCs.

To select candidate variables for this measure, we began with the list of 42 administrative claims-based risk-adjustment variables included in the current publicly reported Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure. To this, we added the NIH Stroke Scale score, which was pulled from the GWTG-Stroke Registry dataset. The registry NIH Stroke Scale data were used as a surrogate for data that will eventually come from claims once ICD-10 codes for stroke severity are available and consistently used by hospitals. Our set of candidate variables from the claims included 2 demographic variables (age and gender), 39 CC-based variables, an indicator variable for when a patient transferred into the hospital from the emergency department, and the NIH Stroke Scale. The final risk-adjustment variables were selected by a team of clinicians and analysts primarily based on their clinical relevance but with knowledge of their strength of association with the mortality outcome.

To develop the model, we began with the 43 candidate predictors for the 30-day mortality and selected the best model using the logistic regression model with the stepwise selection method based on 1,000 bootstrapping samples for each copy of the multiple imputed data. Variable selection rate for all the variables selected into the best model was calculated for each copy of the multiple imputed data, and variables were included into the final model if the minimum variable selection rate among the 5 copies of multiple imputed datasets was 90% or more.

The clinician team reviewed these results and they decided to retain all other risk-adjustment variables above a 90% cutoff, since they demonstrated a relatively strong association with mortality and were clinically relevant (20 variables).

This resulted in a final risk-adjustment model that included 20 variables (see Section 2b4.4a table of candidate variables).

Socioeconomic Factors and Race

We selected variables representing socioeconomic (SES) factors and race for examination based on a review of literature, conceptual pathways, and feasibility. In Section 1.8, we describe the variables that we considered and analyzed based on this review. Below we describe the pathways by which SDS and race may influence 30-day mortality.

Our conceptualization of the pathways by which patient SES or race affects 30-day mortality is informed by the literature.

Literature Review of Socioeconomic (SES) and Race Variables and Stroke Mortality

To examine the relationship between SES and race variables and hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following stroke hospitalization, a literature search was performed with the following exclusion criteria: articles published more than 10 years ago, international studies, articles using Veterans Affairs databases as the primary data source, and articles not explicitly focused on SES or race and stroke mortality. After applying the first exclusion criterion, 86 articles were reviewed by title and abstract, and 77 were excluded from full-text review. Nine articles were selected for full-text review. Among articles reviewed, several found an increased risk of in-hospital mortality associated with SES and/or race variables (Ovbiagele et al., 2010; Hasan et al., 2010), and others found an overall increase in mortality risk not confined to the inpatient hospitalization (Khan et al., 2011; Pedigo et al., 2011; Glymour et al., 2009; Clark et al., 2011; Boan et al., 2014, Howard et al., 2011). Some results showed no increased mortality risk associated with race/ethnicity but a risk conferred by SES (Hanchate et al., 2013).

Causal Pathways for Socioeconomic (SES) and Race Variable Selection

Although some recent literature evaluates the relationship between patient SES or race and the mortality outcome, few studies directly address causal pathways or examine the role of the hospital in these pathways. Moreover, the current literature examines a wide range of conditions and risk variables with no clear consensus on which risk factors demonstrate the strongest relationship with mortality. The SES factors that have been examined in the mortality literature can be categorized into three domains: (1) patient-level variables, (2) neighborhood/community-level variables, and (3) hospital-level variables. Patient-level variables describe characteristics of individual patients, and range from the self-reported or documented race or ethnicity of the patient to the patient's income or education level (Alter et al., 2014; Taksler et al., 2012). Neighborhood/community-level variables use information from sources such as the American Community Survey (ACS) as either a proxy for individual patient-level data or to measure environmental factors. Studies using these variables use one dimensional measures such as median household income or composite

measures such as the Agency for Healthcare Research and Quality (AHRQ)-validated SES index score (Blum et al., 2014). Hospital-level variables measure attributes of the hospital which may be related to patient risk. Examples of hospital-level variables used in studies are ZIP code characteristics aggregated to the hospital level or the proportion of Medicaid patients served in the hospital.

The conceptual relationship, or potential causal pathways by which these possible SES risk factors influence the risk of mortality following an acute illness or major surgery, like the factors themselves, are varied and complex. There are at least four potential pathways that are important to consider.

1. **Relationship of socioeconomic (SES) factors or race to health at admission**. Patients who have lower income/education/literacy or unstable housing may have a worse general health status and may present for their hospitalization or procedure with a greater severity of underlying illness. These SES risk factors, which are characterized by patient-level or neighborhood/community-level (as proxy for patient-level) variables, may contribute to worse health status at admission due to competing priorities (restrictions based on job, lack of childcare), lack of access to care (geographic, cultural, or financial), or lack of health insurance. Given that these risk factors all lead to worse general health status, this causal pathway should be largely accounted for by current clinical risk-adjustment**.**

In addition to SES risk factors, studies have shown that worse health status is more prevalent among African-American patients compared with white patients. The association between race and worse health is in part mediated by the association between race and SES risk factors such as poverty or disparate access to care associated with poverty or neighborhood. The association between race and worse health status that is independent of poverty is mediated through bias in healthcare as well as in other facets of society.

2. **Use of low-quality hospitals**. Patients of lower income, lower education, or unstable housing have been shown not to have equitable access to high quality facilities because such facilities are less likely to be found in geographic areas with large populations of poor patients; thus patients with low income are more likely to be seen in lower quality hospitals, which can contribute to increased risk of mortality following hospitalization (Jha et al., 2011; Reames et al., 2014). Similarly African-American patients have been shown to have less access to high quality facilities compared with white patients (Skinner et al., 2005).

3. **Differential care within a hospital**. The third major pathway by which SES factors or race may contribute to mortality risk is that patients may not receive equivalent care within a facility. For example, African-American patients have been shown to experience differential, lower quality, or discriminatory care within a given facility (Trivedi et al., 2014). Alternatively, patients with SES risk factors such as lower education may require differentiated care – for example, provision of lower literacy information – that they do not receive.

4. **Influence of SES on mortality risk outside of hospital quality and health status**. Some SES risk factors, such as income or wealth, may affect the likelihood of mortality without directly affecting health status at admission or the quality of care received during the hospital stay. For instance, while a hospital may make appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-discharge due to competing economic priorities or a lack of access to care outside of the hospital.

These proposed pathways are too complex to distinguish analytically. They also have different implications on the decision to risk adjust or not. We, therefore, first assessed if there was evidence of a meaningful effect on the risk model to warrant efforts to distinguish among these pathways. Based on this model and the considerations outlined in Section 1.8, the following SES and race variables were considered:

- African American race (as compared to all others)
- Dual eligible status
- AHRQ SES index score

We assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable model. For this measure, we also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results. Given no meaningful improvement in the risk-model or change in performance scores we did not further seek to distinguish the causal pathways for these measures.

References:

Alter DA, Franklin B, Ko DT, et al. Socioeconomic status, functional recovery, and long-term mortality among patients surviving acute myocardial infarction. PloS one 2014; 8:e65130.

Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. Circulation Cardiovascular quality and outcomes 2014; 7:391-7.

Boan AD, Feng WW, Ovbiagele B, et al. Persistent racial disparity in stroke hospitalization and economic impact in young adults in the buckle of stroke belt. Stroke; a journal of cerebral circulation. Jul 2014;45(7):1932-1938.

Clark CJ, Guo H, Lunos S, et al. Neighborhood cohesion is associated with reduced risk of stroke mortality. Stroke; a journal of cerebral circulation. May 2011;42(5):1212-1217.

Glymour MM, Kosheleva A, Boden-Albala B. Birth and adult residence in the Stroke Belt independently predict stroke mortality. Neurology. Dec 1 2009;73(22):1858-1865.

Hanchate AD, Schwamm LH, Huang W, Hylek EM. Comparison of ischemic stroke outcomes and patient and hospital characteristics by race/ethnicity and socioeconomic status. Stroke; a journal of cerebral circulation. Feb 2013;44(2):469-476.

Hasan O, Orav EJ, Hicks LS. Insurance status and hospital care for myocardial infarction, stroke, and pneumonia. Journal of hospital medicine. Oct 2010;5(8):452-459.

Howard VJ, Kleindorfer DO, Judd SE, et al. Disparities in stroke incidence contributing to disparities in stroke mortality. Ann Neurol 2011;69:619–627.

Jha AK, Orav EJ, Epstein AM. Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and medicaid patients. Health affairs 2011; 30:1904-11.

Khan JA, Casper M, Asimos AW, et al. Geographic and sociodemographic disparities in drive times to Joint Commission-certified primary stroke centers in North Carolina, South Carolina, and Georgia. Preventing chronic disease. Jul 2011;8(4):A79.

Krumholz H, Normand S, Galusha D, et al. Risk-Adjustment Methodology for Hospital Monitoring/Surveillance and Public Reporting Supplement #1: 30-Day Mortality Model for Pneumonia. 2006. Available at: https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830. Accessed November 19, 2015.

Normand S-LT, Shahian DM. Statistical and Clinical Aspects of Hospital Outcomes Profiling. 2007/05 2007:206-226.

Ovbiagele B. Nationwide trends in in-hospital mortality among patients with stroke. Stroke; a journal of cerebral circulation. Aug 2010;41(8):1748-1754.

Pedigo A, Seaver W, Odoi A. Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: fuzzy cluster and discriminant analyses approaches. PloS one. 2011;6(7):e22693.

Reames BN, Birkmeyer NJ, Dimick JB, Ghaferi AA. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.

Skinner J, Chandra A, Staiger D, Lee J, McClellan M. Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. Circulation 2005; 112:2634-41.

Taksler GB, Keating NL, Cutler DM. Explaining racial differences in prostate cancer mortality. Cancer 2012; 118:4280-9.

Trivedi AN, Nsa W, Hausmann LR, et al. Quality and equity of care in U.S. hospitals. The New England journal of medicine 2014; 371:2298-308.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Below are tables showing the candidate variables and the final variables that were included more than 90% of the time for all copies of the imputed data and therefore retained in the final model.

Candidate Variables

Variable	Code(s)
NIH Stroke Scale score (continuous)	n/a
Age minus 65 (years above 65, continuous)	n/a
Male	n/a
Transfer from another ED	n/a
Congestive heart failure	CC 80
Valvular or rheumatic heart disease	CC 86
Congenital cardiac/circulatory defects	CC 87-88
Hypertensive heart disease	CC 90
Specified arrhythmias	CC 92
Cerebral hemorrhage	CC 95
Ischemic or unspecified stroke	CC 96
Precerebral arterial occlusion and transient cerebral ischemia	CC 97
Cerebral atherosclerosis and aneurysm	CC 98
Hemiplegia/hemiparesis	CC 100
History of infection	CC 1, 3-6
Metastatic cancer, acute leukemia and other severe cancers	CC 7-8
Lymphatic, head and neck, brain, and other major cancers; breast, colorectal and other major cancers	CC 9-13
Protein-calorie malnutrition	CC 21
Other significant endocrine and metabolic disorders	CC 22-24
Other gastrointestinal disorders	CC 36
Disorders of the vertebrae and spinal discs	CC 39
Osteoarthritis of hip or knee	CC 40
Other musculoskeletal and connective tissue disorders	CC 43
Iron deficiency or other unspecified anemias and blood disease	CC 47
Dementia or other specified brain disorders	CC 49-50
Major psychiatric disorders	CC 54-56
Quadriplegia, other extensive paralysis	CC 67-69
Multiple sclerosis	CC 72, 76
Seizure disorders and convulsions	CC 74
Hypertension	CC 89, 91
Vascular disease and complications	CC 104-105
Chronic obstructive pulmonary disease (COPD)	CC 108
Pneumonia	CC 111-113
Pleural effusion/pneumothorax	CC 114
Other eye disorders	CC 124
Other ear, nose, throat, and mouth disorders	CC 127
Dialysis status	CC 130

Variable	Code(s)
Renal failure	CC 131
Urinary tract infection	CC 135
Male genital disorders	CC 140
Decubitus ulcer of skin	CC 148
Chronic ulcer of skin, except decubitus	CC 149
Other dermatological disorders	CC 153

Final Model Variables (variables meeting criteria in field 2b4.3)

Variable	Code(s)	07/2011 – 06/2014 OR (95% CI)
Age (continuous, per 5 units)	--	1.34 (1.32, 1.36)
Transfer from another ED	--	1.35 (1.26, 1.44)
NIH Stroke Scale score (continuous, per 5 units)	--	1.59 (1.57, 1.61)
Congestive heart failure	CC 80	1.24 (1.17, 1.3)
Congenital cardiac/circulatory defects	CC 87-88	0.66 (0.57, 0.77)
Specified heart arrhythmias	CC 92	1.35 (1.29, 1.41)
Cerebral atherosclerosis and aneurysm	CC 98	0.81 (0.76, 0.87)
Metastatic cancer and acute leukemia and other major cancers	CC 7-8	2.89 (2.63, 3.16)
Protein - calorie malnutrition	CC 21	1.61 (1.5, 1.73)
Other significant endocrine and metabolic disorders	CC 22-24	0.70 (0.66, 0.74)
Other gastrointestinal disorders	CC 36	0.90 (0.86, 0.94)
Disorders of the vertebrae and spinal discs	CC 39	0.88 (0.83, 0.93)
Osteoarthritis of hip or knee	CC 40	0.87 (0.81, 0.93)
Other musculoskeletal and connective tissue disorders	CC 43	0.90 (0.85, 0.94)
Iron deficiency and other/unspecified anemia and blood disease	CC 47	1.19 (1.14, 1.25)
Dementia or other specified brain disorders	CC 49-50	1.26 (1.2, 1.32)
Multiple sclerosis	CC 72, 76	0.87 (0.81, 0.93)
Seizure disorders and convulsions	CC 74	1.23 (1.14, 1.32)
Pneumonia	CC 111-113	1.31 (1.24, 1.38)
Renal failure	CC 131	1.14 (1.08, 1.2)

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Variation in prevalence of the factor across measured entities

The prevalence of SES factors and African-American patients in the Stroke cohort varies across hospitals. The median percentage of dual eligible patients is 9.5% (interquartile range [IQR] 4.1%-15.9%). The median percentage of African-American patients is 4.2% (IQR 0.0%-14.3%). The median percentage of patients with an AHRQ SES index score equal to or below 47.4 is 20.0% (IQR 3.6%-45.0%).

Empirical association with the outcome (univariate)

The patient-level observed stroke mortality rate is somewhat higher for dual eligible patients, 15.5%, compared with 14.1% for all other patients. The mortality rate for African-American patients was lower at 10.0% compared with 14.8% for patients of all other races. The mortality rate for patients with an AHRQ SES index score equal to or below 47.4 was slightly lower at 14.0% compared with 14.4% for patients with an AHRQ SES index score above 47.4.

Incremental effect of SES variables and race in a multivariable model

When we included these variables in a multivariate model that included all of the claims-based clinical variables, only African American race was significantly associated with a lower risk of mortality, with an odds ratio of 0.62. Neither dual eligible or low AHRQ SES index were significant in the multivariable model. In all cases the c-statistics for the stroke patient-level multivariate models with the SDS variables in the models were essentially unchanged from those without (model with original variables: 0.8176; model with dual eligible variable: 0.8176; model with race variable: 0.8184; model with AHRQ SES index variable: 0.8176).

To further understand the relative importance of these risk-factors in the measure we compared hospital performance with and without the addition of each SDS variable. We find that the addition of any of these variables into the model had little to no effect on hospital performance. The mean absolute change in hospitals' RSMRs when adding a dual eligibility indicator was 0.00006% with a correlation coefficient between RSMRs for each hospital with and without dual eligibility added of 0.9999. The mean absolute change in hospitals' RSMRs when adding a race indicator was -0.00064% with a correlation coefficient between RSMRs for each hospital with and without race added of 0.9906. The mean absolute change in hospitals' RSMRs when adding a low SES AHRQ indicator was 0.00009% with a correlation coefficient between RSMRs for each hospital with and without low SES added of 0.9999.

Overall we find that the SES variables that could be feasibly incorporated into this model do not have a significant relationship with the outcome in multivariable modeling. For race the relationship with mortality was in the opposite direction than what has been the expressed concern of stakeholders interested in adding such adjustment to the models. We also find that the impact of any of these indicators is very small to negligible on model performance and hospital profiling. Given the controversial nature of incorporating such variables into a risk-model we do not support doing so in a case that is unlikely to affect hospital profiling.

Given these findings and complex pathways that could explain any relationship between SES or race and mortality, which do not all support risk-adjustment, we did not incorporate SES variables and race into the measure.

Future reevaluation efforts will explore the relationship between SES or race and stroke once ICD-10 data are available.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Approach to assessing model performance

During measure development, we computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the development and validation cohort:

Discrimination Statistics:

(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome.

(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk patients from low-risk patients. Therefore, we would hope to see a wide range between the lowest decile and highest decile)

(3) R-squared indicates how well data fit a statistical model, or the percent of variance explained by the model.

Calibration Statistics:

(4) Over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients).

We tested the performance of the model developed in a randomly selected 50% sample of the hospitalizations for ischemic stroke in **Dataset 1** (development dataset; July 2011-June 2014) by comparing results with those from the validation sample (the remaining 50% of the dataset).

References:

Harrell FE, Shih YCT. Using full probability models to compute probabilities of actual interest to decision makers. *Int. J. Technol. Assess. Health Care* 17 (2001), pp. 17–26.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

For the development cohort (**Dataset 1**), the results are summarized below:
• 1st half of randomly split sample (development sample): C-statistic = 0.8124; Predictive ability (lowest decile %, highest decile %) = (1.33, 50.00); Adjusted R-squared = 0.2681
• 2nd half of randomly split sample (validation sample): C-statistic = 0.8165; Predictive ability (lowest decile %, highest decile %) = (1.26, 51.57); Adjusted R-squared = 0.2764

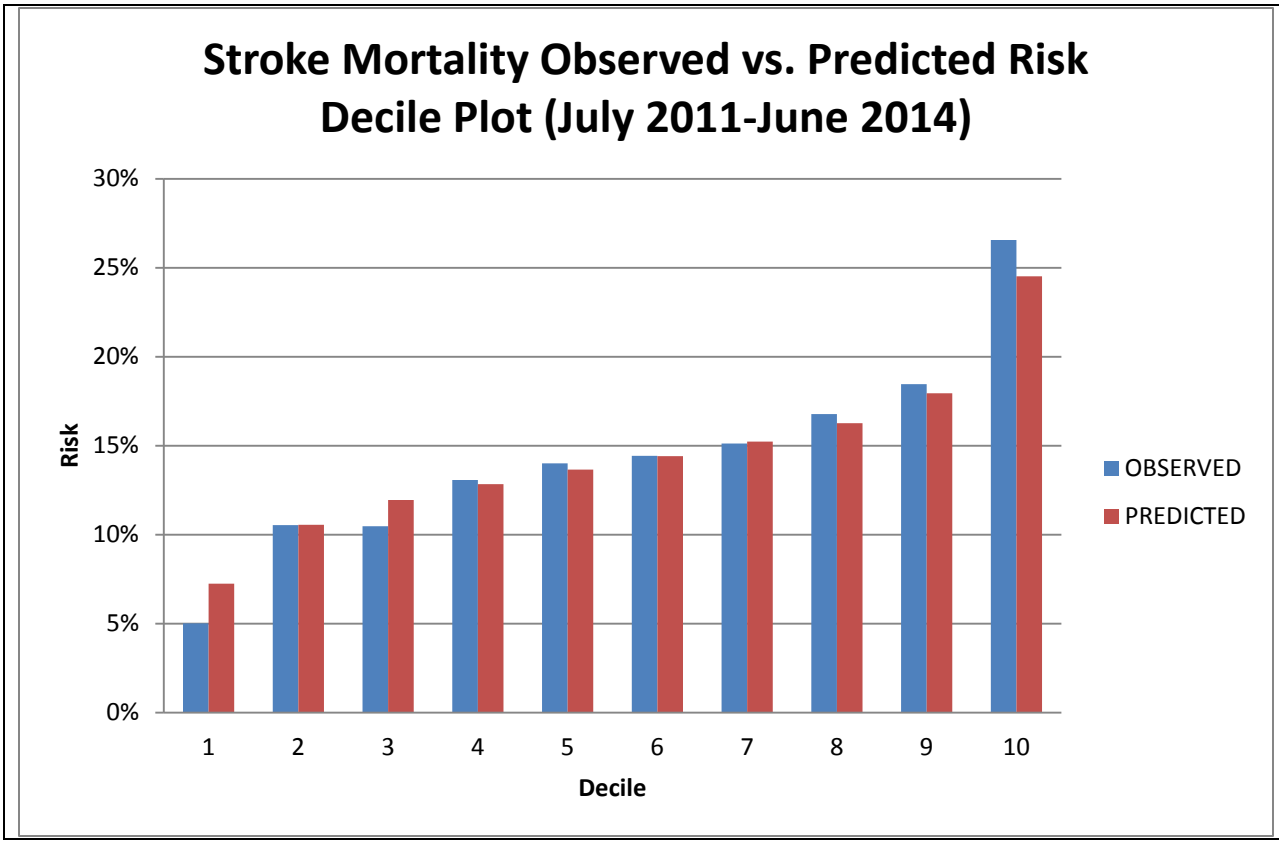
For comparison of model with and without inclusion of SES factors, see above section.

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

For the development cohort (**Dataset 1**), the results are summarized below:
• 1st half of randomly split sample (development sample) Calibration: (0.0000, 1.0000)
• 2nd half of randomly split sample (validation sample) Calibration: (0.0000, 1.0000)

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

The risk decile plot is a graphical depiction of the observed mortality in the deciles of the predicted mortality to measure predictive ability. Below, we present the risk decile plot showing the distributions for the development dataset (**Dataset 1**). The plot for the validation dataset was similar.



2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Discrimination Statistics

The C-statistics of 0.8124 in the development sample and 0.8165 in the validation sample indicate good model discrimination (**Dataset 1**). The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics

Over-fitting (Calibration γ0, γ1)

If the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration values of almost 0 at one end and almost 1 on the other end indicate good calibration of the model.

Risk Decile Plots

Higher deciles of the predicted outcomes are associated with higher observed outcomes, which show a good calibration of the model. This plot indicates excellent discrimination of the model and good predictive ability.

Overall Interpretation

Interpreted together, our diagnostic results demonstrate the risk-adjustment model adequately controls for differences in patient characteristics (case mix).

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

N/A

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The method for discriminating hospital performance has not been determined. For public reporting of measures of hospital outcomes developed with similar methodology, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSMR's interval estimate does not include the national observed mortality rate (is lower or higher than the rate), then CMS is confident that the hospital's RSMR is different from the national rate, and describes the hospital on the Hospital Compare website as "better than the U.S. national rate" or "worse than the U.S. national rate." If the interval includes the national rate, then CMS describes the hospital's RSMR as "no different than the U.S. national rate" or "the difference is uncertain." CMS does not classify performance for hospitals that have fewer than 25 cases in the three-year period.

However, the measure is not currently publicly reported and decisions about the approach to discriminating hospital performance have not been made.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Analyses of Medicare FFS data show substantial variation in RSMRs among hospitals (**Dataset 1**). Using data from the development sample (**Dataset 1**, July 2011-June 2014), the median hospital RSMR was 14.29%, with a range of 11.27% to 17.54%. The interquartile range was 13.63% - 15.05%.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

Despite recent decreases in mortality rates nationally, stroke is the fifth most common cause of death in the United States, affecting approximately 795,000 people annually, and has a 30-day mortality rate that varies by age from 9% in patients 65 to 74 years of age, 13.1% in those 74 to 84 years of age, and 23% in those ≥85 years of age [Go et al., 2014; Kochanek et al., 2014; Casper et al., 2008].

The variation in RSMRs suggests that there are differences in the quality of care received across hospitals for stroke that support measurement to reduce this variation.

References:

Casper ML, Nwaise IA, Croft JB, Nilasena DS. Atlas of Stroke Hospitalizations Among Medicare Beneficiaries. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention; 2008.

Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. *Circulation.* Jan 21 2014;129(3):e28-e292.

Kochanek KD MS, Xu JQ, Arias E. Mortality in the United States, 2013. NCHS data brief, no 178. 2014.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

There is no current national database that includes NIH Stroke Scale score data for stroke patients admitted to all non-federal acute care hospitals. Therefore, implementation of this measure depends on hospitals including patients' NIH Stroke Scale scores, for all patients admitted with ischemic stroke, in the claims they submit to Medicare using the ICD-10 codes that will be available beginning in October 2016. Once these ICD-10 codes become available, it will be important to assess rates of missing data. Collection of the NIH Stroke Scale is now Class I recommended in the AHA/ASA guidelines for care of patients admitted with acute ischemic stroke. Given the importance of this measure in national reporting programs and the importance of stroke severity to this measure's risk model, CMS expects high rates of NIH Stroke Scale reporting and will determine the plan for handling missing data during measure implementation.

GWTG-Stroke Registry Data Element: NIH Stroke Scale Score

For measure development we assessed the frequency of missing NIH Stroke Scale scores from the GWTG-Stroke Registry during the development of the measure. The registry data on stroke severity was used as a surrogate for data that will eventually come from the claims once the ICD-10 codes for stroke severity are available and consistently used by hospitals. Among the measure cohort, NIH Stroke Scale scores were available in 82.39% of patients with an admission for ischemic stroke from July 1, 2013 and June 30, 2014. This proportion increased from 70.78% of admissions occurring between July 1, 2011 and June 30, 2012. Using the registry stroke severity data, we used multiple imputation to produce

estimates for our testing and analyses. The multiple imputation technique used to impute missing values was a multi-logit regression model. Five copies of imputation datasets were produced for the analyses. The results based on these data were aggregated according to the standard statistical methods for the presentation of the results and for the measure score calculation.

In multiple imputation, missing variable values are predicted using other related patient variables available. The predicted values are substituted for the missing values, which results in a full data set without any missing variables (the imputed data set). By repeating this process multiple times, we get multiple imputed data sets. We then conduct analyses on and obtain results for each imputed data set. The results based on multiple data sets are combined to produce the overall final results. The multiple imputation represents a random sample of the missing values according to the association of the non-missing values of all the variables considered, and the resulting inferences of multiple imputation are statistically valid, which reflect the uncertainty due to missing values (He et al., 2014).

To determine if patterns of missing stroke severity data were systematic among hospitals participating in the GWTG-Stroke Registry in a way that might bias measure results, we examined the correlation between the rate of missing NIH Stroke Scale among patients and hospital-level 30-day mortality rate for all admissions, those with NIH Stroke Scale scores and those missing NIH Stroke Scale scores. We also examined the correlation among only those admissions with missing NIH Stroke Scale scores. While this statistical measure was developed using multiple imputation to account for missing NIH Stroke Scale data, approaches to handling missing NIH Stroke Scale data in measure calculation will be reassessed during implementation.

References:

He R, Belin T. Multiple imputation for high-dimensional mixed incomplete continuous and binary data. Stat. Med. 2014;33:2251–2262.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

In the development dataset which used registry data as a surrogate for future claims data on stroke severity, NIH Stroke Scale scores were available in 82.39% of patients with an admission for ischemic stroke from July 1, 2013 and June 30, 2014. This proportion increased from 70.78% of admissions occurring between July 1, 2011 and June 30, 2012. For all 3 years of data combined, the NIH Stroke Scale score was missing in 23.51% of all patients (N=188,975). At the hospital level, the median percentage of missing NIH Stroke Scale scores was 10.15% (IQR: 3.89%-27.27%), with a range of 0% to 100%, between July 1, 2013 and June 30, 2014. This percentage improved from July 1, 2011 to June 30, 2012, when the median percentage of missing NIH Stroke Scale scores was 19.23% (IQR: 7.15%-50.00%), with a range of 0% to 100%. For all 3 years of data combined, the median percentage of missing NIH Stroke Scale scores at the hospital level was 15.89% with a range of 0% to 100%. The interquartile range was 6.74% to 39.27%.

The correlation between the rate of missing NIH Stroke Scale scores in the registry dataset and the hospital-level 30-day mortality rate was not significant overall. In the subset of patients who had missing NIH Stroke Scale scores, the correlation between the rate of missing NIH Stroke Scale scores and the hospital-level 30-day mortality rate was -0.2568 ($p<0.0001$).

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

The overall rate of missing NIH Stroke Scale was relatively low. The results are not biased due to systematic missing data (or differences between patients with reported versus unreported data), as we used multiple imputation to substitute missing values with predicted values. The proportion of patients missing NIH Stroke Scale scores had little impact on hospital 30-day mortality rate. However, the stroke severity data from the GWTG-Stroke Registry are only used as a surrogate for data that will eventually come from Medicare claims. These analyses will need to be repeated in the future using a claims dataset.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*

ALL data elements are in defined fields in electronic claims

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.

Administrative data are routinely collected as part of the billing process. New ICD-10 codes for NIH Stroke Scale scores will be available to hospitals to include in Medicare claims in October 2016.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.

There are no fees associated with the use of this measure.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the

time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Not in use	

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

This is a new measure.

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

CMS intends to implement this measure in the Hospital Inpatient Quality Reporting (Hospital IQR) Program once the new NIH Stroke Scale ICD-10 codes have been in use for 3 years. This measure requires 3 years of claims data for calculation. This timeline depends on the consistent capture of NIH Stroke Scale scores and use of the new ICD-10 codes by hospitals for all patients admitted with acute ischemic stroke as is recommended in the current AHS/ASA clinical guidelines. Once this new measure is implemented it will replace the currently reported Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

This is a new measure and there is no information available on performance improvement. However, there has been significant progress in 30-day RSMR for stroke when looking at the current publicly reported measure. The median 30-day RSMR decreased by 1.1 absolute percentage points from 2011-2012 (median RSMR: 15.3%) to 2013-2014 (median RSMR: 14.2%). The median hospital RSMR from 2011-2014 was 14.9% (IQR 14.2% - 15.6%).

4b.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Once this measure is ready for public reporting, CMS intends to replace the stroke measure that is currently publicly reported. This new stroke measure has improved credibility and face validity among stakeholders and increased the ability to differentiate hospital performance using NIH Stroke Scale.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such

evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**
We did not identify any unintended consequences during measure development or model testing. However, we are committed to monitoring this measure's use and assessing potential unintended consequences over time, such as the inappropriate shifting of care, increased patient morbidity and mortality, and other negative unintended consequences for patients.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**
0467 : Acute Stroke Mortality Rate (IQI 17)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**
Title: Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure
Steward: Centers for Medicare & Medicaid Services

5a. Harmonization

The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications completely harmonized?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
We did not include in our list of related measures any non-outcome (such as process) measures with the same target population as our measure. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure). Additionally, this measure and the NQF endorsed Acute Stroke Mortality Rate (IQI 17) (AHRQ) Measure #0467 are complementary and related rather than competing measures. Although they both assess mortality for patients admitted to acute care hospitals with a principal discharge diagnosis of acute ischemic stroke, the specified outcomes are different. Our measure assesses 30-day mortality, while #0467 assesses inpatient mortality. The 30-day mortality and inpatient mortality outcomes each have distinct advantages and uses, which make them complementary (and related) as opposed to competing. For example the 30-day period provides a broader perspective on hospital care and utilizes a standard time period to examine hospital performance to avoid bias by differences in length of stay among hospitals. However, in some settings it may not be feasible to capture post-discharge mortality, making the inpatient measure more useable. We have previously consulted with AHRQ to examine harmonization of the measures' cohort. As a result of that collaboration, we have found that the measures' cohorts are harmonized to the extent possible and that the small differences in cohort inclusion and exclusion criteria are appropriate because the measures assess different outcomes.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
This measure looks at a longer outcome time frame (30-days versus in-hospital) and incorporates stroke severity into the risk-model.

The current publicly reported measure, Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure, is a potentially competing measure. It is CMS intent to replace the current measure in any given program with this newly developed measure, which includes stroke severity in the risk model.

The Hybrid Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Acute Ischemic Stroke with Risk Adjustment for Stroke Severity measure is also being submitted to NQF for endorsement. This measure uses a combination of claims and electronic health records (EHR) data for risk adjustment but is otherwise harmonized with the new claims-only measure. It is CMS intent to implement only one of the new stroke mortality measures (this claims-only measure or the hybrid measure) in any given program.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Claims-Only_-_Hybrid_Stroke_Mortality_Measure_Tech_Report_1-15-16-635884734566322294.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
Co.2 **Point of Contact:** Lein, Han, Lein.han@cms.hhs.gov, 410-786-0205-
Co.3 **Measure Developer if different from Measure Steward:** Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)
Co.4 **Point of Contact:** Karen, Dorsey, karen.dorsey@yale.edu, 203-764-5700-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Our working group consisted of the following members:
- Lee Schwamm, MD: Vice Chairman, Department of Neurology, Massachusetts General Hospital
- Gregg Fonarow, MD: Professor of Medicine, University of California, Los Angeles
- Jason Sico, MD: Director, Stroke Care VA Connecticut Healthcare System
- Kevin Sheth, MD: Associate Professor of Neurology and Neurosurgery, Yale University

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?** N/A
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:** N/A
Ad.7 **Disclaimers:** N/A

Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 2877
Measure Title: Hybrid hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute ischemic stroke with risk adjustment for stroke severity
Measure Steward: Centers for Medicare & Medicaid Services (CMS)
Brief Description of Measure: This hybrid stroke mortality measure estimates the hospital-level, risk-standardized mortality rate (RSMR) for patients discharged from the hospital with a principal discharge diagnosis of acute ischemic stroke. The outcome is all-cause 30-day mortality, defined as death from any cause within 30 days of the index admission date, including in-hospital death, for stroke patients. This measure is a newly developed measure with a cohort and outcome that is harmonized with the CMS's current publicly reported claims-based stroke mortality measure, and includes the National Institutes of Health (NIH) Stroke Scale as an assessment of stroke severity in the risk-adjustment model. The measure is referred to as a hybrid because it is CMS's intention to calculate the measure using two data sources: Medicare fee-for-service (FFS) administrative claims and clinical electronic health record (EHR) data.
Developer Rationale: Stroke is the fifth most common cause of death, affecting approximately 795,000 people in the United States annually, and has a mortality rate of 17% [Go et al., 2014; Kochanek et al., 2014]. Stroke is also a leading cause of disability in the United States, which can lead to increased dependency on the health care system and higher subsequent costs associated with this care [Centers for Disease Control and Prevention, 2005]. Mortality following stroke – an important adverse outcome that can be measured reliably and objectively, and that is influenced by the quality of care provided to patients during their initial hospitalization – is an appropriate measure of quality of care [DesHarnais et al., 1988; Weir et al, 2001]. Specifically, post-stroke mortality rates have been shown to be influenced by critical aspects of care such as response to complications, speediness of delivery of care, organization of care, and appropriate imaging [Smith et al., 2006; Reeves et al., 2009; Lingsma et al., 2008; Hong et al., 2008]. This work demonstrates the relationship between hospital organizational factors and performance on the stroke mortality measure and supports the ability of hospitals to impact these rates.

The goal of outcome measurement is to identify institutions whose performance is better or worse than would be expected based on their patient case mix by risk-adjusting for patients' conditions at the time of hospital admission. The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level RSMRs following hospitalization for acute ischemic stroke. Measurement of patient mortality allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures.

Rationale for Development of a Hybrid Stroke Mortality Measure

Current outcome measures use administrative claims data from the year prior to the index admission in the risk adjustment models. Stakeholders, including the AHA/ASA and other professional organizations, have indicated a preference for clinical data over administrative claims data in the risk adjustment models of stroke mortality measures.

They have specifically highlighted the importance of including stroke severity in mortality measures for risk adjustment, as several studies have demonstrated that initial stroke severity is the strongest predictor of mortality in ischemic stroke patients [Smith et al., 2010; Nedeltchev et al., 2010; Fonarow et al., 2012].

This hybrid stroke mortality measure addresses these stakeholder preferences and improves measure performance by incorporating clinical data and stroke severity scores into the risk-adjustment model. The recent proliferation of EHR systems, standardization of data extraction for quality reporting, and advancements in clinical practice to incorporate new clinical assessments have made it possible to integrate these data into measures of hospital performance. For example, the NIH Stroke Scale, which was created in 1989, is guideline-endorsed and widely used in routine stroke care. Collection of the NIH Stroke Scale is required at over 1,700 hospitals participating in the AHA/ASA GWTG -Stroke throughout the U.S. [Fonarow et al., 2014]. Utilization of these data will not only address stakeholder preferences, but will also improve the discrimination of the risk models.

References:

1. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. Circulation. Jan 21 2014;129(3):e28-e292.
2. Kochanek KD MS, Xu JQ, Arias E. Mortality in the United States, 2013. NCHS data brief, no 178. 2014.
3. Centers for Disease Control and Prevention (CD). Prevalence and most common causes of disability among adults: United States, 2005. MMWR Morb Mortal Wkly Rep. 2009;58:421-426.
4. Casper ML NI, Croft JB, Nilasena DS. Atlas of Stroke Hospitalizations Among Medicare Beneficiaries. 2008.
5. DesHarnais SI, Chesney JD, Wroblewski RT, Fleming ST, McMahon LF, Jr. The Risk-Adjusted Mortality Index. A new measure of hospital performance. Med Care. Dec 1988;26(12):1129-1148.
6. Weir NU, Sandercock PA, Lewis SC, Signorini DF, Warlow CP. Variations between countries in outcome after stroke in the International Stroke Trial (IST). Stroke. Jun 2001;32(6):1370-1377.
7. Smith MA, Liou JI, Frytak JR, Finch MD. 30-day survival and rehospitalization for stroke patients according to physician specialty. Cerebrovascular diseases (Basel, Switzerland). 2006;22(1):21-26.
8. Reeves MJ, Smith E, Fonarow G, Hernandez A, Pan W, Schwamm LH. Off-hour admission and in-hospital stroke case fatality in the get with the guidelines-stroke program. Stroke. Feb 2009;40(2):569-576.
9. Lingsma HF, Dippel DW, Hoeks SE, et al. Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey. Journal of neurology, neurosurgery, and psychiatry. Aug 2008;79(8):888-894.
10. Hong KS, Kang DW, Koo JS, et al. Impact of neurological and medical complications on 3-month outcomes in acute ischaemic stroke. European journal of neurology : the official journal of the European Federation of Neurological Societies. Dec 2008;15(12):1324-1331.
11.Smith EE, Shobha N, Dai D, et al. Risk score for in-hospital ischemic stroke mortality derived and validated within the Get With the Guidelines-Stroke Program. Circulation. 2010;122(15):1496-1504.
12. Nedeltchev K, Renz N, Karameshev A, et al. Predictors of early mortality after acute ischaemic stroke. Swiss Medical Weekly. 2010;140(17-18):254-259.
13. Fonarow GC, Saver JL, Smith EE, et al. Relationship of national institutes of health stroke scale to 30-day mortality in medicare beneficiaries with acute ischemic stroke. J Am Heart Assoc. Feb 2012;1(1):42-50.
14. Adams HP, Jr., Davis PH, Leira EC, et al. Baseline NIH Stroke Scale score strongly predicts outcome after stroke: A report of the Trial of Org 10172 in Acute Stroke Treatment (TOAST).
15. Fonarow GC, Alberts MJ, Broderick JP, et al. Stroke outcomes measures must be appropriately risk adjusted to ensure quality care of patients: a presidential advisory from the American Heart Association/American Stroke Association. Stroke. May 2014;45(5):1589-1601.

Numerator Statement: The outcome for this measure is 30-day, all-cause mortality. We define mortality as death from any cause within 30 days of the index admission for patients with a principal discharge diagnosis of acute ischemic

stroke.
Denominator Statement: The cohort includes inpatient admissions for Medicare FFS patients, age 65 years and older, who were discharged from non-federal, short-term, acute care hospitals with a principal discharge diagnosis of acute ischemic stroke.

Additional details are provided in S.9 Denominator Details.
Denominator Exclusions: The measure excludes admissions for patients:
1. With inconsistent or unknown vital status or other unreliable data;
2. Enrolled in the Medicare hospice program at any time in the 12 months prior to the index admission, including the first day of the index admission; and
3. Discharged against medical advice (AMA).
For patients with more than one admission for stroke in a given year, only one index admission for that condition is randomly selected for inclusion in the cohort.

Measure Type: Outcome
Data Source: Administrative claims, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Registry, Other
Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Summary of evidence:

- This is a newly-developed hybrid measure that calculates hospitals' 30-day risk-standardized mortality rate for patients who have been hospitalized with an ischemic stroke. This measure is harmonized with the CMS's current publicly reported claims-based stroke mortality measure and includes the National Institutes of Health (NIH) Stroke Scale as an assessment of stroke severity in the risk-adjustment model. In 2012, the Neurology Steering Committee considered a similar measure that did not include the NIHSS in the risk-adjustment approach but the Committee could not come to consensus about the measure and it was withdrawn from consideration by the developer.
- As a rationale for measuring this health outcome, the developer states notes that many hospital processes have been associated with lower stroke mortality rates within 30 days of hospital admission including prevention of, and response to, complications, speediness of delivery of care, organization of care, appropriate imaging, patient safety, and coordinated transitions to the outpatient environment.
- The developer reports studies demonstrate appropriate, guideline-recommended care and timely treatment for stroke patients can reduce the risk of mortality within 30 days of hospital admission.

Question for the Committee:

- *Does the SC agree that at least one hospital process identified by the developer impacts ischemic stroke*

mortality rates within 30 days of admission?

Guidance from the Evidence Algorithm : Health outcome (Box 1) →relationship between outcome and at least one healthcare action identified/supported by stated rationale (Box 2) → Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

Per the developer, they combined two data sources (for model development purposes only): July 2011 – June 2014 Medicare Administrative claims and 2013 AHA/ASA GWTG-Stroke Registry. Registry data served as surrogate for those data elements that can be feasibly extracted from EHRs, including the NIH Stroke Scale score.

- **Sample size**: 188,975 patients at 1,511 hospitals
- **Mean** risk-standardized mortality rate (RSMR): 14.55%
- **Median** risk-standardized mortality rate (RSMR): 14.50%
- **Range**: 10.67% to 19.15%
- **Interquartile range**: 13.50% to 15.61%

The developer also provided the results of CMS's current publicly reported claims-based stroke mortality measure (this is a different measure based only on data from administrative claims; it does not include the NIH Stroke Scale in the risk-adjustment model) as reported in 2014 update to Hospital Compare which demonstrated variation across hospitals.

- **Sample size:** 520,111 admissions from 4,506 hospitals
- **Median** risk-standardized mortality rate (RSMR): 15.3%
- **Interquartile range**: 8.6% to 23.8%

Disparities

- The developer provides the following information (July 2011-June 2014):

	#hospitals	*# admissions*	*Minimum rate*	*Median rate*	*Maximum rate*
Dual eligibles					
Low proportion	551	19,859	11.56	14.24	17.73
High proportion	294	14,376	11.33	14.20	17.60
African Americans					
Low proportion	492	13,497	12.04	14.25	17.73
High proportion	294	17,429	11.56	14.13	17.11
AHRQ SES score					
Low proportion	295	11,563	11.71	14.25	16.74
High proportion	294	22,187	11.33	14.16	16.97

Questions for the Committee:

- *Does this gap in care warrant a national performance measure?*
- *Are you aware of evidence that disparities exist in this area of healthcare?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus

Comments: **yes. there is a dire need to link outcomes specifically mortality to severity of patient clinical presentation

**The outcome of mortality is certainly a strong outcome to measure. I appreciate the exclusion of patients with hospice up to the day of admission since there are outcomes worse than death however I worry that this will still penalize hospitals that are better and helping patients and families focus on QOL versus Length of life.

1b. Performance Gap

Comments: **there is gap data, but the measure appears to be more applicable to all ischemic stroke patients

**A performance gap was demonstrated by variability in performance.

I do not think the data based on race and SES demonstrated disparities in care

1c. High Priority (previously referred to as High Impact)

Comments: **yes

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): The developer listed administrative claims, electronic clinical data, electronic health record, laboratory data, and registry as the data sources. However, this measure is a "hybrid" measure that is specified for EHRs and administrative claims only.

Specifications:

- The measure is specified as a facility-level measure for the hospital/acute care setting.
- The denominator includes all Medicare FFS beneficiaries, age 65 and over, with a principal discharge diagnosis of acute ischemic stroke.
    - During the 2012 NQF evaluation of the stroke mortality measure (which relied only on administrative claims data), in an effort to harmonize with other mortality/readmissions measures, the developer included all-payer patients ages 18 and over (rather than Medicare FFS patient ages 65+ only). Other mortality measures can be used in either of two patient cohorts: (1) patients 65 years or older, or (2) patients 18 years or older.
- ICD-9 and ICD-10 codes provided; ICD-9 to ICD-10 crosswalk available in the data dictionary found in S.2b.
- Exclusions include inconsistent or unknown vital status, enrolled in Medicare hospice any time in the 12 months prior to the index admission including first day of admission, and discharged AMA; exclusion details provided.
- This outcome measure is risk-adjusted using a statistical risk-adjustment model with 21 factors including age, patient clinical data, and multiple comorbidities.
    - In 2013, "ED transfer" was added to the risk-adjustment model in the current publically reported claims-based stroke mortality measure.
    - This updated version of the measure (hybrid) includes the NIH Stroke Scale in the risk-adjustment model—this data element could be obtained from either claims or EHR systems. NIHSS values will be available via ICD-10 dx code beginning October 2016.
- The calculation algorithm, included in S.18, describes how the risk-standardized mortality ratio is calculated.
- HQMF specifications for the EHR component of this hybrid measure are included in the document set on

SharePoint. See eMeasure Technical Review below.

Questions for the Committee :

- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

eMeasure Technical Advisor review

Submitted measure is an HQMF compliant eMeasure	The submitted eMeasure specifications follow the industry accepted format for eMeasure (HL7 Health Quality Measures Format (HQMF)). **HQMF specifications** ☒ **Yes** ☐ **No**
Documentation of HQMF or QDM limitations	Submitted eMeasure contains components that cannot be represented due to limitations of HQMF or QDM and the submission explains the work around for these limitations. This is a hybrid measure using both claims data and data elements found in electronic health records.
Value Sets	The submitted eMeasure specifications uses existing value sets when possible and uses new value sets that have been vetted through the VSAC .
Measure logic is unambiguous	Submission includes test results from a simulated data set demonstrating the measure logic can be interpreted precisely and unambiguously.
Feasibility Testing	The submission contains a feasibility assessment that addresses data element feasibility and follow-up with measure developer indicates that the measure logic is feasible based on assessment by EHR vendors.

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☐ **Yes** ☒ **No**

Method(s) of reliability testing:

- Dataset used for testing included Medicare Parts A and B claims, Medicare Enrollment Database(EDB), and AHA/ASA Get With The Guidelines (GWTG) Stroke Registry. The registry data were used as a surrogate for data that will eventually come from electronic health records (EHRs), or, in the case of stroke severity, from claims once ICD-10 codes for stroke severity are available and consistently used by hospitals. Reliable coding is expected but CMS will monitor implementation of the new ICD-10 codes for NIH Stroke Scale scores available October 2016.
- A subset of 1,520 non-federal, acute inpatient US hospitals that participate in GWTG Stroke Registry were included in the dataset.
  - For measure development purposes only, the developers linked two data sources: Medicare claims and GWTG Stroke Registry data. Registry data were used to obtain the NIH Stroke Scale scores and clinical risk variables that could in the future be extracted from EHRs.
  - The developers then merged Medicare Part A Inpatient, Outpatient, and Part B Outpatient claims with

GWTG Stroke Registry stroke severity and clinical data from July 1, 2011 to June 30, 2014 which resulted in the following data sample (**Dataset 1**):

- Number of hospitals: 1,511
- Number of total admissions: 188,975
- First half of split sample (development sample): 94,466 admissions; 1,473 hospitals
- Second half of split sample (validation sample): 94,509 admissions; 1,462 hospitals

- Although the developer reported assessing data element reliability by comparing model variable frequencies and odds ratios from logistic regression models across the most recent three years of data, NQF does not consider temporal consistency to be a valid method of demonstrating reliability of data elements.

- Developers used a split-sample (or "test-retest") methodology to test the measure score reliability; this is an appropriate method. Developers randomly assigned half of the patients in each hospital to two separate groups, calculated the performance measure score for each hospital in each of the two groups, and compared the agreement between each hospital's paired scores using the intra-class-correlation coefficient (ICC) and applying a correction factor to account for the overall sample size. The ICC reflects the percentage of variance in score results that is due to "true" or real variance between the hospitals.
- Data element reliability of electronic clinical data elements: data element validity testing was performed and will count for data element reliability as well – see validity testing section

Results of reliability testing:

- Measure score reliability results:
  - The ICC values from the split-sample analysis was 0.56, indicating 56% of the variance in scores are due to differences between hospitals. According to the Landis and Koch classification, an ICC value of 56% can be interpreted as moderate agreement. However, a value of 0.7 is often regarded as a minimum acceptable reliability value.
  - The developer notes that this analysis was limited to hospitals with 12 or more cases in each split sample although the measure is not specified to include a minimum data sample of 12 cases. The developer also notes that the ICC is based on a split sample of three years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data – the measure is intended to be reported with three years of data.

Questions for the Committee:

- *Do the testing results presented by the developer demonstrate an adequate level of reliability?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Assessments of performance score reliability often examine the ability of the measure to differentiate between measured entities. Do the reliability testing results reported by the developer demonstrate that meaningful differences in performance can be identified?*

Algorithm 2 Reliability for claims-based data elements: Precise specifications (Box 1) →empiric reliability testing (Box 2) →performance score testing (Box 4) →appropriate method of testing (Box 5) → moderate certainty of reliability (Box 6b)

Algorithm 2 Reliability for electronic clinical data elements: Precise specifications (Box 1) →empiric reliability testing (Box 2 → empiric validity testing of patient-level data (Box 3) → YES: Use rating from validity testing of patient-level data elements →3 patient-level clinical data elements validity testing conducted (Box 10)→EHR abstraction comparted to gold standard (Box 11)→High or moderate certainty that data used in the measure are valid (Box 12a)→Moderate

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

- This measure estimates 30-day all-cause mortality for patients hospitalized with ischemic stroke using a risk standardized mortality ratio (RSMR), which is calculated as the ratio of the number of "predicted" to the number of "expected" deaths, multiplied by the national unadjusted mortality rate.
- As a rationale for measuring this health outcome, the developers suggest that hospitals are able to influence mortality rates through a broad range of clinical activities, including prevention of, and response to, complications, speediness of delivery of care, organization of care, appropriate imaging, patient safety, and coordinated transitions to the outpatient environment.

Question for the Committee:

o *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

☐ **Face validity only**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- Data element validity of electronic clinical data elements: The developer validated all 3 electronically abstracted critical EHR data elements (heart rate, diastolic blood pressure, and glucose) against manual chart abstraction which is considered the gold standard.
- The developer states that face validity was systematically assessed by outside experts and the public throughout the development of this measure. The outside experts formed a working group which consisted of various expertise to ensure the measure is meaningful, useful, and well-designed. The developers also solicited public comments through a 30-day public comment period. NQF guidance requires that face validity of the measure score must explicitly address whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. It does not appear that the face validity assessment conducted by the developer conforms to NQF's requirements.
- The developer conducted empirical validity testing of the measure score by comparing this measure to two similar stroke mortality measures. All 3 models include a total of 188,975 hospital admissions derived from registry and claims data (**Dataset 1**).
    - An updated claims-only risk-adjustment model that includes only variables obtained from administrative claims data (demographics, comorbidities, and patient medical history) and the NIH Stroke Scale score (which can be captured using administrative claims data starting in October 2016). Because the NIHSS is not yet available from claims, this data element was obtained from the GWTG-Stroke Registry.
    - A clinical-only risk-adjustment model that includes only variables from electronic clinical data (demographics, laboratory results, and vital signs), as well as the NIH Stroke Scale score (which can be

captured using electronic clinical data).

- The developer noted that each of the three cohorts for the three risk models used the same inclusion/exclusion criteria and a risk-adjustment (statistical modeling) strategy and only differed with respect to the risk variables used.

Validity testing results:

- Data element validity results:
  - Table 3b demonstrates the percent agreement between electronically abstracted critical EHR data elements to that obtained from manual chart abstraction in two different EHR systems:

Data Element/ CCDE	% Agreement Between Datasets (Number Matching/ Total Records With A Data Value)
Dataset 4 (n=368)	
Heart rate (BPM)	95.55 (322/337)
Diast Blood Pressure (mmHG)	94.38 (319/338)
Glucose (mg/dL)	96.14 (274/285)
Dataset 5 (n=391)	
Heart rate (BPM)	57.45 (135/235)
Diast Blood Pressure (mmHG)	60.09 (137/228)
Glucose (mg/dL)	95.12 (78/82)

  - The developer states that extracting errors in Dataset 5 that were not present in Dataset 4, drove down the accuracy of the two data elements, heart rate and diastolic blood pressure.
- The performance of the claims-only risk-adjustment model and the clinical-only risk-adjustment model was similar to the performance of the hybrid risk-adjustment model . The areas under the receiver operating characteristic (ROC) curve (or c-statistic) were **0.81** and **0.79**, respectively, for the two models compared with **0.82** for the hybrid risk-adjustment model.
  - A c-statistic is a model of discrimination statistic. A c-statistic of 0.82 means that 82% of all possible pairs of patients – one who died and one who lived – the model correctly assigned a higher probability to those who died. Generally, a c-statistic of at least 0.70 is considered acceptable. The similar c-statistics indicate that all 3 models have similar discriminatory ability.
- The correlation coefficient between the RSMRs from the claims-only risk-adjustment model and the hybrid risk-adjustment model was 0.97. The correlation coefficient between the RSMRs from the clinical-only risk-adjustment model and the hybrid risk-adjustment model was 0.96.
- The developer states that the empirical validity testing results between the hybrid stroke model and the other two models suggest that each approach shows similarly high discrimination and the c-statistic was "nearly identical". The developer also concluded that the high correlation among the RSMRs calculated from the 3 models shows that each model provides similar/consistent results for hospitals.

Questions for the Committee:

- *In the absence of NIHSS values from ICD-10 codes, is use of values from the GWTG-Stroke registry a reasonable surrogate?*
- *Do the methods and results demonstrate sufficient validity so that conclusions about quality can be made?*

- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- To determine impact of exclusions on the cohort (188,975 hospital admissions), the developer examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion:
  - Inconsistent or unknown vital status or other unreliable data – **0.0%**
  - Enrolled in the Medicare hospice program at any time in the 12 months prior to the index admission, including the first day of the index admission – **0.78%**
  - Discharged against medical advice (AMA) – **0.22%**
- Exclusions 1 and 2 are necessary to calculate the measure; exclusion 3 is needed for acceptability of the measure to hospitals who do not have the opportunity to deliver full care and prepare the patient for discharge.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

2b4. Risk adjustment: **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary:

Description of the model

- This measure is risk-adjusted using hierarchical logistic regression model with **21 factors** to create a hospital-level 30-day risk-standardized mortality rate that simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals.
  - The model adjusts for case differences based on age, transfer from another ED, NIH Stroke Scale score, and clinical status of the patient at the time of admission (the latter using condition categories (CCs)).
  - Only comorbidities that conveyed information about the patient at the time of admission or in the 12-months prior, and not complications that arose during the course of the hospitalization, were included in the risk-adjustment.
- To select candidate variables for this measure the developers began with the list of 42 variables included in the current publicly reported Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure and added the NIH Stroke Scale score and 14 clinical data elements in the GWTG-Stroke Registry considered feasibly extractable from the EHR. The final risk-adjustment model included 21 variables that demonstrated a relatively strong association with mortality and were clinically relevant.

Performance of the model

Discrimination statistics:

- The *c*-statistic reflects how accurately a statistical model is able to distinguish between a patient with an outcome and a patient without an outcome. C-statistic values can range from 0.5 to 1.0. A value of 0.5 indicates that the model is no better than chance at making a prediction of patients with and without the outcome of interest and a value of 1.0 indicates that the model perfectly identifies those with and without the outcome of interest . Generally, a c-statistic of at least 0.70 is considered acceptable.
- The *c*-statistic value was computed using data from **Dataset 1** which was randomly split into two samples:

  - The *c*-statistic for the 1st half of the randomly split sample (development sample) was **0.8176**.
  - The *c*-statistic for the 2nd half of the randomly split sample (validation sample) was **0.8206**.
- A *c*-statistic of **0.8176** means that for 81.7% of all possible pairs of patients—one who died and one who lived—the model correctly assigned a higher probability to those who died. The similar c-statistic in the validation model shows that the risk model has similar discriminatory ability in a dataset different from the one from which it was developed.
- Developers also noted that the model indicated a wide range between the lowest decile and highest decile which indicated the ability to distinguish between high-risk patients and low-risk patients.
  - 1st half of randomly split sample (development sample): **1.31%, 51.05%**
  - 2nd half of randomly split sample (validation sample): **1.05%, 51.93%**

Calibration statistics:

- The developers noted that the risk-decile plot based on the development dataset demonstrated excellent discrimination and good predictive ability of the risk-adjustment model because the observed values are relatively similar to the predicted values across the risk-deciles. The developers also noted that the plot for the validation dataset demonstrated similar results.
- According to the developers, the calibration values of almost 0 and almost 1, for the development sample and the validation sample, indicated good calibration of the model (i.e., the model is not overfit).

Conceptual basis and empirical support for potential inclusion of SDS factors in risk-adjustment approach

- The developer noted that although some recent literature has evaluated the relationship between patient SES or race and mortality, few studies directly address causal pathways or examine the role of the hospital in these pathways. Additionally, there is no clear consensus in the literature on which risk factors demonstrate the strongest relationship with mortality. They note SES factors that have been examined in the mortality literature include: patient-level self-reported or documented race or ethnicity, income, and education level; occupational level; median household income; Agency for Healthcare Research and Quality (AHRQ)-validated SES index score; and the proportion of Medicaid patients served in the hospital.

The developer identified several potential conceptual pathways to consider:

  - Relationship of socioeconomic status (SES) factors or race to health at admission.
  - Use of low-quality hospitals.
  - Differential care within a hospital.
  - Influence of socioeconomic status (SES) on mortality risk outside of hospital quality and health status.
- Based on their interpretation of the literature and analysis of the above pathways, the developers identified 3 potential SDS variables for potential inclusion in the risk-adjustment model:
  - African American race (as compared to all others)
  - Dual eligible status
  - AHRQ SES index score (based on 5-digit ZIP code data; includes percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th-grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room)
- Analyses indicate that the prevalence of these 3 SDS factors varies across hospitals and are associated with the measured outcome. Of the 3 SDS factors, only African American race was significantly associated with a lower risk of mortality (OR=0.62, which indicates a protective effect). The developers note that the c-statistics for the models that included the individual SDS factors along with the original variables were similar (original=0.8176; with dual eligible=0.8176; with race= 0.8184; with AHRQ SES index=0.8176).

Addition of these factors individually resulted in very little change in the RSMRs (e.g., for race, the average absolute change in hospitals' RSMRs was -0.00064%).

- Based on these results, the developer decided **NOT** to include any of the 3 SDS factors analyzed in the final risk-adjustment model.

Questions for the Committee:

- *Does the risk model adequately control for differences in case mix across providers?*
- *Are all of the risk adjustment variables present at the start of care? If not, describe the rationale provided.*

Do you agree with the developer's decision, based on their analysis, to not include SDS factors in their risk-adjustment model?

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developers provided statistical results from the dataset used to develop this measure (Dataset 1, July 2011-July 2014) which demonstrated substantial variation in RSMRs among hospitals. The median hospital RSMR was 14.29% with a range of 11.27% to 17.54%. The interquartile range was 13.63% - 15.05%.The developers provided additional mortality rates from the literature.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

- The developer describes a second approach to risk adjustment for this measure that uses only clinical data elements from patient EHRs and provided the results of the analysis of the performance of that alternate risk model. They also compare the performance of all three risk models, the hybrid model, the EHR-only alternate model, and the claims-only model, in 2b2.3 and 2b2.4.

2b7. Missing Data

- The developers assessed the frequency of missing NIH Stroke Scale scores from the GWTG Stroke Registry during the development of this measure. The registry data were used as a surrogate for data that will eventually come from claims once ICD-10 codes for stroke severity are available and consistently used by hospitals. The analysis performed on the development dataset found:
  - July 1, 2011 – June 30, 2012: NIH Stroke Scale scores available in **70.78%** admissions for ischemic stroke
  - July 1, 2013 – June 30, 2014: NIH Stroke Scale scores available in **82.39%** of ischemic stroke admissions
- The developers used multiple imputation to substitute missing values of stroke severity with predicted values and found that the proportion of patients missing NIH Stroke Scale scores had little impact on hospital 30-day mortality rates. However, these analyses will need to be repeated in the future using a claims dataset.
- Due to the importance of this measure in national reporting programs and the importance of stroke severity to this measure's risk model the developer states that CMS will determine the plan for handling missing data during the measure implementation.

Guidance from Validity algorithm: Precise specifications (Box 1) → potential threats to validity assessed (Box 2) → empirical validity testing (Box 3) →measure score testing (Box 6) → method conceptually sound conceptualization (Box 7) → moderate confidence that scores are a valid indicator of quality (Box 8b)

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

2a1. & 2b1. Specifications

Comments: **clinical evidence suggests that the severity of the presenting ischemic stroke does influence long term outcome **Some of the citations cited by the developers specifically state that the Quality of Care has limited impact on mortality which is not what they describe in their summary.

2a2. Reliability Testing

Comments: **have concerns regarding reliability among hospitals

2b2. Validity Testing

Comments: **no

2b3. Exclusions Analysis

2b4. Risk Adjustment/Stratification for Outcome or Resource Use Measures

2b5. Identification of Statistically Significant & Meaningful Differences In Performance

2b6. Comparability of Performance Scores When More Than One Set of Specifications

2b7. Missing Data Analysis and Minimizing Bias

Comments: **yes

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- All data elements are in defined fields in a combination of electronic sources and generated or collected by and used by healthcare personnel during the provision of care. The data are coded by someone other than the person obtaining original information.
- Administrative data are routinely collected as part of the billing process. New ICD-10 codes for NIH Stroke Scale scores will be available to hospitals to include in Medicare claims in October 2016.
- Electronic clinical data will be collected from hospitals using MAT output and value sets to inform data queries and electronic reporting requirements.
- The data element feasibility assessment scorecard submitted by the developer demonstrated a feasibility score of "3" (highest rating) for the required clinical data elements (heart rate, diastolic blood pressure, and glucose) on all four components.
    - Data availability was tested in 3 separate health systems and 3 EHRs (Epic, Cerner, and GE Centricity).
    - Data accuracy was tested in 2 hospitals and 2 EHRs (Cerner and GE Centricity).

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*
- *Does the eMeasure Feasibility Score Card demonstrate acceptable feasibility in multiple EHR systems and sites?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

3a. Byproduct of Care Processes

3b. Electronic Sources

3c. Data Collection Strategy

Comments: **n/a

**This measure (hybrid) relies on the NIH stroke score in the electronic record and ICD 10 coding and I am not sure that we have the data we need to know how reliable that data is going to be.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details:

- CMS intends to implement this measure in the Hospital Inpatient Quality Reporting (Hospital IQR) Program once the new NIH Stroke Scale ICD-10 codes and the core clinical data elements (CCDE) have been in use for 3 years. This measure requires 3 years of claims data for calculation. Once one of the new measures (either the claims-based or hybrid measure) is implemented it will replace the currently reported Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure.

Potential harms:

- The developer did not identify any unintended consequences related to this measure.

Feedback :

- The measure currently in use, "Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure" was evaluated by NQF in 2012. The lack of inclusion of the NIH Stroke Scale was a major topic during public comment and in committee discussion. The committee was split regarding recommending the measure without including the NIH Stroke Scale in the risk-adjustment model. The measure was withdrawn from consideration by the measure developer.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

4a. Accountability and Transparency

4b. Improvement

4c. Unintended Consequences

Comments: **need to incorporate NIHSS into reimbursement mechanism

Criterion 5: Related and Competing Measures

Related or competing measures:

- 0467 : Acute Stroke Mortality Rate (IQI 17)
- 2876 : Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Acute Ischemic Stroke with Claims-Based Risk Adjustment for Stroke Severity
- Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure (not NQF endorsed)

Harmonization:

- The developers stated that measures' cohorts are harmonized to the extent possible and that the small differences in cohort inclusion and exclusion criteria are appropriate because the measures (0467, 2876, & 2877) assess different stroke outcomes – inpatient mortality vs. 30-day mortality.
- In 2012 when the original 30-day stroke mortality measure was submitted and reviewed, as part of their harmonization efforts with other similar mortality measures, the developer re-specified the measure to include all-payer patients ages 18 and over (rather than Medicare FFS patients ages 65+ only). *This updated version of the measure **does not** include all-payer patients ages 18 and over.*
- 2876 is also being submitted to NQF for endorsement; this measure uses administrative claims data for risk adjustment. The developer noted that 2876 is otherwise harmonized with this new hybrid measure. It is CMS's intent to implement only one of the new stroke mortality measures (this hybrid measure or the claims-only measure) in any given program.
- It is CMS's intent to replace the current publicly reported measure, Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization, in any given program with this newly developed measure, which includes stroke severity in the risk model.

Pre-meeting public and member comments

No public comments were submitted for this measure during the pre-meeting commenting period.

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Hybrid hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute ischemic stroke hospitalization with risk adjustment for stroke severity
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: Click here to enter a date

Instructions

- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- Respond to all questions as instructed with answers immediately following the question. All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 10 pages (*incudes questions/instructions*; minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating

Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome

☒ Health outcome: Hybrid hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute ischemic stroke hospitalization with risk adjustment for stroke severity
Health outcome includes patient-reported outcomes (PRO, i.e., HRQoL/functional status, symptom/burden, experience with care, health-related behaviors)

☐ Intermediate clinical outcome: Click here to name the intermediate outcome

☐ Process: Click here to name the process

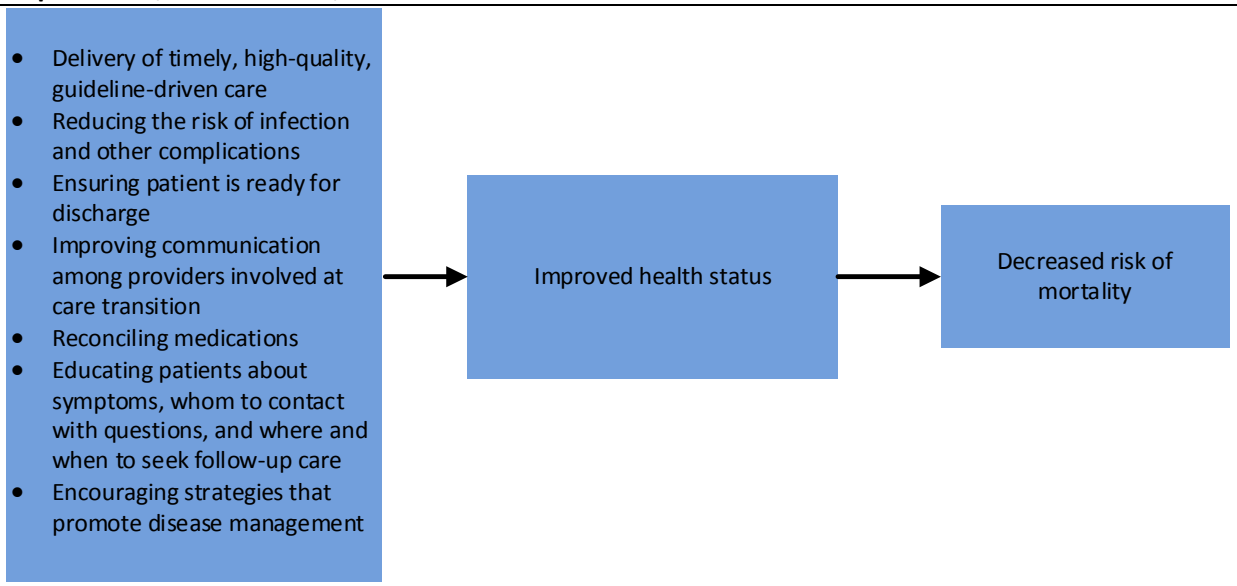
☐ Structure: Click here to name the structure

☐ Other: Click here to name what is being measured

_____

HEALTH OUTCOME/PRO PERFORMANCE MEASURE *If not a health outcome or PRO, skip to 1a.3*

1a.2. Briefly state or diagram the path between the health outcome (or PRO) and the healthcare structures, processes, interventions, or services that influence it.



The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level, risk-standardized mortality rates following hospitalization for stroke. Measurement of patient outcomes allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures. Complex and critical aspects of care, such as communication between providers, prevention of, and response to, complications, patient safety and coordinated transitions to the outpatient environment, all contribute to patient outcomes but are difficult to measure by individual process measures. The goal of outcomes measurement is to risk-adjust for patients' conditions at the time of hospital admission and then evaluate patient outcomes. This mortality measure was developed to identify institutions, whose performance is better or worse than would be expected based on their patient case-mix, and therefore promote hospital quality improvement and better inform consumers about care quality.

1a.2.1. State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process, intervention, or service (*i.e., influence on outcome/PRO*).

Stroke is a priority condition for outcomes measure development because it is a leading cause of morbidity and mortality. Stroke affects as many as 795,000 individuals in the United States each year and is the nation's fifth leading cause of death (CDC-NCHS, 2015). It is estimated that stroke costs $34 billion each year in direct and indirect medical costs (Mozaffarian et al., 2015).

Many current hospital processes have been associated with lower stroke mortality rates within 30 days of hospital admission. In particular, post-stroke mortality rates have been shown to be influenced by critical aspects of care at the hospital such as response to complications, speediness of delivery of care, organization of care, and appropriate imaging (Smith et al., 2006; Reeves et al., 2009; Lingsma et al., 2008; Hong et al., 2008; Fonarow et al., 2014). This research demonstrates the relationship between hospital organizational factors and performance on the stroke mortality measure, and supports the ability of hospitals to impact these rates. Stakeholders have previously highlighted the importance of including stroke severity in mortality measures for risk adjustment, as several studies have demonstrated that initial stroke severity is the strongest predictor of mortality in ischemic stroke patients (Smith et al., 2010; Nedeltchev et al., 2010; Fonarow et al., 2012). This update to the current publicly reported measure responds to stakeholder preference to include the National Institutes of Health (NIH) Stroke Scale as an assessment of stroke severity and clinical data element in the risk-adjustment model, thereby accounting for stroke severity at the time of admission to assess the condition of the patient before care has been administered.

Complex and critical aspects of care – such as communication between providers, prevention of and response to complications, patient safety, and coordinated transitions to the outpatient environment – all contribute to patient outcomes. For example, hospitals participating in quality improvement registries like Get With The Guidelines (GWTG) had lower in-hospital mortality rates among stroke patients than hospitals not participating in similar programs (Fonarow et al., 2014). Risk-adjusted measures of patient outcomes, specifically mortality, can highlight variations in the provision of care, and thus support improvements by highlighting institutions that provide exceptional care for stroke patients.

The diagram above indicates some of the many care processes that can influence mortality risk. Numerous studies have demonstrated that appropriate (guideline recommended care) and timely treatment for stroke patients can reduce the risk of mortality within 30 days of hospital admission (Hacke et al., 2004; Fang et al., 2008).

The one cost for hospitals associated with this measure is the requirement to collect and report the clinical data elements to CMS for centralized calculation of the measure.

References:

CDC, NCHS. Underlying Cause of Death 1999-2013 on CDC WONDER Online Database, released 2015. Data are from the Multiple Cause of Death Files, 1999-2013, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed Feb. 3, 2015.

Fang J, Keenan NL, Ayala C, Dai S, Merritt R, Denny CH. Awareness of stroke warning symptoms—13 states and the District of Columbia, 2005. MMWR. 2008;57(18):481–5.

Fonarow GC, Saver JL, Smith EE, et al. Relationship of national institutes of health stroke scale to 30-day mortality in medicare beneficiaries with acute ischemic stroke. J Am Heart Assoc. Feb 2012;1(1):42-50.

Fonarow GC, Zhao X, Smith EE, et al. Door-to-Needle Times for Tissue Plasminogen Activator Administration and Clinical Outcomes in Acute Ischemic Stroke Before and After a Quality Improvement Initiative. JAMA. 2014;311(16):1632-1640.

Hacke W, Donnan G, Fieschi C, et al. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. Lancet (London, England). Mar 6 2004;363(9411):768-774.

Hong KS, Kang DW, Koo JS, et al. Impact of neurological and medical complications on 3-month outcomes in acute ischaemic stroke. European journal of neurology : the official journal of the European Federation of Neurological Societies. Dec 2008;15(12):1324-1331.

Lingsma HF, Dippel DW, Hoeks SE, et al. Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey. Journal of neurology, neurosurgery, and psychiatry. Aug 2008;79(8):888-894.

Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. Circulation. 2015 ;e29-322.

Nedeltchev K, Renz N, Karameshev A, et al. Predictors of early mortality after acute ischaemic stroke. Swiss Medical Weekly. 2010;140(17-18):254-259.

Reeves MJ, Smith E, Fonarow G, Hernandez A, Pan W, Schwamm LH. Off-hour admission and in-hospital stroke case fatality in the get with the guidelines-stroke program. Stroke. Feb 2009;40(2):569-576.

Smith EE, Shobha N, Dai D, et al. Risk score for in-hospital ischemic stroke mortality derived andvalidated within the Get With the Guidelines-Stroke Program. Circulation. 2010;122(15):1496-1504.

Smith MA, Liou JI, Frytak JR, Finch MD. 30-day survival and rehospitalization for stroke patients according to physician specialty. Cerebrovascular diseases (Basel, Switzerland). 2006;22(1):21-26.5.

Note: For health outcome/PRO performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURE

1a.3. Briefly state or diagram the path between structure, process, intermediate outcome, and health outcomes. Include all the steps between the measure focus and the health outcome.

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

1a.3.1. What is the source of the systematic review of the body of evidence that supports the performance measure?

☐ Clinical Practice Guideline recommendation – ***complete sections 1a.4, and 1a.7***
☐ US Preventive Services Task Force Recommendation – ***complete sections 1a.5 and 1a.7***
☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*) – ***complete sections 1a.6 and 1a.7***
☐ Other – ***complete section 1a.8***

N/A. This measure is not an intermediate outcome, process, or structure performance measure.

Please complete the sections indicated above for the source of evidence. You may skip the sections that do not apply.

1a.4. CLINICAL PRACTICE GUIDELINE RECOMMENDATION

1a.4.1. Guideline citation (*including date*) and **URL for guideline** (*if available online*):

N/A

1a.4.2. Identify guideline recommendation number and/or page number and **quote verbatim, the specific guideline recommendation**.

N/A

1a.4.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.4.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: If separate grades for the strength of the evidence, report them in section 1a.7.*)

N/A

1a.4.5. Citation and URL for methodology for grading recommendations (*if different from 1a.4.1*)**:**

N/A

1a.4.6. If guideline is evidence-based (rather than expert opinion), are the details of the quantity, quality, and consistency of the body of evidence available (e.g., evidence tables)?

☐ Yes → ***complete section 1a.7***

☐ No → ***report on another systematic review of the evidence in sections 1a.6 and 1a.7; if another review does not exist, provide what is known from the guideline review of evidence in 1a.7***

N/A

1a.5. UNITED STATES PREVENTIVE SERVICES TASK FORCE RECOMMENDATION

1a.5.1. Recommendation citation (*including date*) and **URL for recommendation** (*if available online*):

N/A

1a.5.2. Identify recommendation number and/or page number and **quote verbatim, the specific recommendation**.

N/A

1a.5.3. Grade assigned to the quoted recommendation with definition of the grade:

N/A

1a.5.4. Provide all other grades and associated definitions for recommendations in the grading system. (*Note: the grading system for the evidence should be reported in section 1a.7.*)

N/A

1a.5.5. Citation and URL for methodology for grading recommendations (*if different from 1a.5.1*)**:**

N/A

Complete section 1a.7

1a.6. OTHER SYSTEMATIC REVIEW OF THE BODY OF EVIDENCE

1a.6.1. Citation (*including date*) and **URL** (*if available online*):

N/A

1a.6.2. Citation and URL for methodology for evidence review and grading (*if different from 1a.6.1*)**:**

N/A

Complete section 1a.7

1a.7. FINDINGS FROM SYSTEMATIC REVIEW OF BODY OF THE EVIDENCE SUPPORTING THE MEASURE

If more than one systematic review of the evidence is identified above, you may choose to summarize the one (or more) for which the best information is available to provide a summary of the quantity, quality, and consistency of the body of evidence. Be sure to identify which review is the basis of the responses in this section and if more than one, provide a separate response for each review.

1a.7.1. What was the specific structure, treatment, intervention, service, or intermediate outcome addressed in the evidence review?

N/A

1a.7.2. Grade assigned for the quality of the quoted evidence with definition of the grade:

N/A

1a.7.3. Provide all other grades and associated definitions for strength of the evidence in the grading system.

N/A

1a.7.4. What is the time period covered by the body of evidence? (*provide the date range, e.g., 1990-2010*). Date range: Click here to enter date range

N/A

QUANTITY AND QUALITY OF BODY OF EVIDENCE

1a.7.5. How many and what type of study designs are included in the body of evidence? (*e.g., 3 randomized controlled trials and 1 observational study*)

N/A

1a.7.6. What is the overall quality of evidence across studies in the body of evidence? (*discuss the certainty or confidence in the estimates of effect particularly in relation to study factors such as design flaws, imprecision due to small numbers, indirectness of studies to the measure focus or target population*)

N/A

ESTIMATES OF BENEFIT AND CONSISTENCY ACROSS STUDIES IN BODY OF EVIDENCE

1a.7.7. What are the estimates of benefit—magnitude and direction of effect on outcome(s) across studies in the body of evidence? (*e.g., ranges of percentages or odds ratios for improvement/ decline across studies, results of meta-analysis, and statistical significance*)

N/A

1a.7.8. What harms were studied and how do they affect the net benefit (benefits over harms)?

N/A

UPDATE TO THE SYSTEMATIC REVIEW(S) OF THE BODY OF EVIDENCE

1a.7.9. If new studies have been conducted since the systematic review of the body of evidence, provide for <u>each</u> new study: 1) citation, 2) description, 3) results, 4) impact on conclusions of systematic review.

N/A

1a.8 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.8.1 What process was used to identify the evidence?

N/A

1a.8.2. Provide the citation and summary for each piece of evidence.

N/A

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
NQF_2877_Hybrid_Stroke_Mortality_NQF_Evidence_Attachment_v1.0.docx

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., the benefits or improvements in quality envisioned by use of this measure)*
Stroke is the fifth most common cause of death, affecting approximately 795,000 people in the United States annually, and has a mortality rate of 17% [Go et al., 2014; Kochanek et al., 2014]. Stroke is also a leading cause of disability in the United States, which can lead to increased dependency on the health care system and higher subsequent costs associated with this care [Centers for Disease Control and Prevention, 2005]. Mortality following stroke – an important adverse outcome that can be measured reliably and objectively, and that is influenced by the quality of care provided to patients during their initial hospitalization – is an appropriate measure of quality of care [DesHarnais et al., 1988; Weir et al, 2001]. Specifically, post-stroke mortality rates have been shown to be influenced by critical aspects of care such as response to complications, speediness of delivery of care, organization of care, and appropriate imaging [Smith et al., 2006; Reeves et al., 2009; Lingsma et al., 2008; Hong et al., 2008]. This work demonstrates the relationship between hospital organizational factors and performance on the stroke mortality measure and supports the ability of hospitals to impact these rates.

The goal of outcome measurement is to identify institutions whose performance is better or worse than would be expected based on their patient case mix by risk-adjusting for patients' conditions at the time of hospital admission. The goal of this measure is to improve patient outcomes by providing patients, physicians, and hospitals with information about hospital-level RSMRs following

hospitalization for acute ischemic stroke. Measurement of patient mortality allows for a broad view of quality of care that encompasses more than what can be captured by individual process-of-care measures.

Rationale for Development of a Hybrid Stroke Mortality Measure

Current outcome measures use administrative claims data from the year prior to the index admission in the risk adjustment models. Stakeholders, including the AHA/ASA and other professional organizations, have indicated a preference for clinical data over administrative claims data in the risk adjustment models of stroke mortality measures. They have specifically highlighted the importance of including stroke severity in mortality measures for risk adjustment, as several studies have demonstrated that initial stroke severity is the strongest predictor of mortality in ischemic stroke patients [Smith et al., 2010; Nedeltchev et al., 2010; Fonarow et al., 2012].

This hybrid stroke mortality measure addresses these stakeholder preferences and improves measure performance by incorporating clinical data and stroke severity scores into the risk-adjustment model. The recent proliferation of EHR systems, standardization of data extraction for quality reporting, and advancements in clinical practice to incorporate new clinical assessments have made it possible to integrate these data into measures of hospital performance. For example, the NIH Stroke Scale, which was created in 1989, is guideline-endorsed and widely used in routine stroke care. Collection of the NIH Stroke Scale is required at over 1,700 hospitals participating in the AHA/ASA GWTG -Stroke throughout the U.S. [Fonarow et al., 2014]. Utilization of these data will not only address stakeholder preferences, but will also improve the discrimination of the risk models.

References:

1. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. Circulation. Jan 21 2014;129(3):e28-e292.
2. Kochanek KD MS, Xu JQ, Arias E. Mortality in the United States, 2013. NCHS data brief, no 178. 2014.
3. Centers for Disease Control and Prevention (CD). Prevalence and most common causes of disability among adults: United States, 2005. MMWR Morb Mortal Wkly Rep. 2009;58:421-426.
4. Casper ML NI, Croft JB, Nilasena DS. Atlas of Stroke Hospitalizations Among Medicare Beneficiaries. 2008.
5. DesHarnais SI, Chesney JD, Wroblewski RT, Fleming ST, McMahon LF, Jr. The Risk-Adjusted Mortality Index. A new measure of hospital performance. Med Care. Dec 1988;26(12):1129-1148.
6. Weir NU, Sandercock PA, Lewis SC, Signorini DF, Warlow CP. Variations between countries in outcome after stroke in the International Stroke Trial (IST). Stroke. Jun 2001;32(6):1370-1377.
7. Smith MA, Liou JI, Frytak JR, Finch MD. 30-day survival and rehospitalization for stroke patients according to physician specialty. Cerebrovascular diseases (Basel, Switzerland). 2006;22(1):21-26.
8. Reeves MJ, Smith E, Fonarow G, Hernandez A, Pan W, Schwamm LH. Off-hour admission and in-hospital stroke case fatality in the get with the guidelines-stroke program. Stroke. Feb 2009;40(2):569-576.
9. Lingsma HF, Dippel DW, Hoeks SE, et al. Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey. Journal of neurology, neurosurgery, and psychiatry. Aug 2008;79(8):888-894.
10. Hong KS, Kang DW, Koo JS, et al. Impact of neurological and medical complications on 3-month outcomes in acute ischaemic stroke. European journal of neurology : the official journal of the European Federation of Neurological Societies. Dec 2008;15(12):1324-1331.
11.Smith EE, Shobha N, Dai D, et al. Risk score for in-hospital ischemic stroke mortality derived and validated within the Get With the Guidelines-Stroke Program. Circulation. 2010;122(15):1496-1504.
12. Nedeltchev K, Renz N, Karameshev A, et al. Predictors of early mortality after acute ischaemic stroke. Swiss Medical Weekly. 2010;140(17-18):254-259.
13. Fonarow GC, Saver JL, Smith EE, et al. Relationship of national institutes of health stroke scale to 30-day mortality in medicare beneficiaries with acute ischemic stroke. J Am Heart Assoc. Feb 2012;1(1):42-50.
14. Adams HP, Jr., Davis PH, Leira EC, et al. Baseline NIH Stroke Scale score strongly predicts outcome after stroke: A report of the Trial of Org 10172 in Acute Stroke Treatment (TOAST).
15. Fonarow GC, Alberts MJ, Broderick JP, et al. Stroke outcomes measures must be appropriately risk adjusted to ensure quality care of patients: a presidential advisory from the American Heart Association/American Stroke Association. Stroke. May 2014;45(5):1589-1601.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for endorsement maintenance. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included). This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
For model development purposes only, we used two data sources: July 2011-June 2014 Medicare Administrative claims and the 2013 AHA/ASA GWTG-Stroke Registry. Both data sources were linked to create the dataset used for measure development. The registry data served as a surrogate for those data elements that can be feasibly extracted from EHR and the NIHSS. Our cohort included 188,975 patients at 1,511 hospitals. The mean risk-standardized mortality rate (RSMR) among hospitals was 14.55%, the median hospital RSMR was 14.50%, with a range of 10.67% to 19.15%, and an interquartile range was 13.50%- to 15.61%).

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
Mortality following stroke is an important adverse outcome that can be measured reliably and objectively. The 30-day stroke mortality rate varies by age from 9% in patients 65-74 years of age to 23% in those =85 years of age [Casper et al., 2008]. Post-stroke mortality rates have also been shown to be influenced by critical aspects of care, as discussed in detail in Section 2.7.

Risk-adjusted measures of patient outcomes, including mortality, can highlight variation in the care patients receive across hospitals, and thus support improvements and learning from high quality institutions. CMS currently publicly reports a claims-based stroke mortality measure. The results of this measure, as reported in the 2014 update to the Hospital Compare, are based on RSMRs calculated for admissions among Medicare FFS patients, age 65 years and older, from July 1, 2010 – June 30, 2013. It includes 520,111 admissions from 4,506 hospitals. The median hospital RSMR was 15.3%, with a range of 8.6% to 23.8%. This variation across hospitals indicates that there is room for improvement in care for stroke patients that could reduce mortality rates.
Additionally, risk-adjusted 30-day mortality rates were shown to decline from 12.1% (95% confidence interval [CI]: 12.0%-12.2%) to 11.6% (95% CI: 11.5%-11.7%) in Medicare FFS from 1999 to 2011 [Krumholz 2014]. This decline suggests that there is opportunity for further improvement in the 30-day mortality outcome over time.

References:
1. Casper ML NI, Croft JB, Nilasena DS. Atlas of Stroke Hospitalizations Among Medicare Beneficiaries. 2008.
2. Krumholz HM, Normand SL-RT, Wang Y. Trends in Hospitalizations and Outcomes for Acute Cardiovascular Disease and Stroke: 1999-2011. 2014.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for endorsement maintenance. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the subcriterion on improvement (4b.1) under Usability and Use.*
Distribution of stroke RSMRs using claims and electronic clinical data for risk adjustment by proportion of Medicaid patients
Dates of data: July 2011 – June 2014
Data source: Medicare FFS claims and AHA registry data

Characteristic//Hospitals with a low proportion (0%) Medicaid patients//Hospitals with a high proportion (>14.2%) Medicaid patients
–
Number of Measured Entities (Hospitals)// 551 // 294
Number of Patients// 19,859 in low-proportion hospitals // 14,376 in high-proportion hospitals
Maximum// 17.73 // 17.60
90th percentile// 15.07 // 15.26
75th percentile// 14.59 //14.69
Median (50th percentile)// 14.24 //14.20
25th percentile// 13.99 // 13.82
10th percentile// 13.58 // 13.36
Minimum // 11.56 // 11.33

Distribution of stroke RSMRs using claims and electronic clinical data for risk adjustment by proportion of African-American patients

Dates of data: July 2011 – June 2014
Data source: Medicare FFS claims and AHA registry data

Characteristic//Hospitals with a low proportion (0%) African-American patients//Hospitals with a high proportion (>17.5%) African-American patients –
Number of Measured Entities (Hospitals)// 492 // 294
Number of Patients// 13,497 in low-proportion hospitals // 17,429 in high-proportion hospitals
Maximum// 17.73 // 17.11
90th percentile// 15.08 // 15.20
75th percentile// 14.60 // 14.65
Median (50th percentile)// 14.25 // 14.13
25th percentile// 14.06 // 13.69
10th percentile// 13.77// 13.14
Minimum // 12.04 // 11.56

Distribution of stroke RSMRs using claims and electronic clinical data for risk adjustment by AHRQ SES Index
Dates of data: July 2011 – June 2014
Data source: Medicare FFS claims and AHA registry data

Characteristic// Hospitals with lowest mean value (=47.68) AHRQ SES index // Hospitals with highest mean value (>59.38) AHRQ SES index Medicaid patients –
Number of Measured Entities (Hospitals)// 295 // 294
Number of Patients// 11,563 in low-proportion hospitals // 22,187 in high-proportion hospitals
Maximum// 16.74 // 16.97
90th percentile// 15.20 //15.35
75th percentile// 14.68 //14.68
Median (50th percentile)// 14.25 // 14.16
25th percentile// 13.95 // 13.62
10th percentile// 13.38// 13.13
Minimum // 11.71 // 11.33

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations.**
N/A

1c. High Priority (previously referred to as High Impact)

The measure addresses:

- a specific national health goal/priority identified by DHHS or the National Priorities Partnership convened by NQF; OR
- a demonstrated high-priority (high-impact) aspect of healthcare (e.g., affects large numbers of patients and/or has a substantial impact for a smaller population; leading cause of morbidity/mortality; high resource use (current and/or future); severity of illness; and severity of patient/societal consequences of poor quality).

1c.1. **Demonstrated high priority aspect of healthcare**
Affects large numbers, A leading cause of morbidity/mortality, High resource use, Patient/societal consequences of poor quality, Severity of illness
1c.2. **If Other:**

1c.3. **Provide epidemiologic or resource use data that demonstrates the measure addresses a high priority aspect of healthcare. List citations in 1c.4.**
Stroke is the fifth most common cause of death in the United States each year with approximately 795,000 people having a stroke annually, and has a mortality rate of 17% [Go et al., 2014]. Stroke is one of CMS's top 10 costliest conditions [Go et al., 2014; Total

Expenses, 2012]. The direct medical costs of stroke in 2010 (which includes hospital outpatient visits, hospital inpatient stays, emergency department visits, prescribed medications, and home health care) was estimated to be $20.6 billion [Total Expenses, 2012].

1c.4. **Citations for data demonstrating high priority provided in 1a.3**

1. Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. Circulation. Jan 21 2014;129(3):e28-e292.
2. Total Expenses and Percent Distribution for Selected Conditions by Source of Payment: United States, 2012. Medical Expenditure Panel Survey Household Component Data. Generated interactively. 2012.

1c.5. **If a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful.** *(Describe how and from whom their input was obtained.)*

N/A. This measure is not a PRO-PM.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*

Neurology, Neurology : Stroke/Transient Ischemic Attack (TIA)

De.6. **Cross Cutting Areas** *(check all the areas that apply):*

Care Coordination, Safety, Safety : Complications

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is an eMeasure **Attachment:** CCDE_v4_Artifacts.zip

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment **Attachment:** NQF_2877_Hybrid_Stroke_Mortality_S2b_Mortality_Data_Dictionary_v1.0.xlsx

S.3. **For endorsement maintenance**, please briefly describe any changes to the measure specifications since last endorsement date and explain the reasons.

N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome)*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.

The outcome for this measure is 30-day, all-cause mortality. We define mortality as death from any cause within 30 days of the index admission for patients with a principal discharge diagnosis of acute ischemic stroke.

S.5. **Time Period for Data** *(What is the time period in which data will be aggregated for the measure, e.g., 12 mo, 3 years, look back to August for flu vaccination? Note if there are different time periods for the numerator and denominator.)*
Numerator time window: We define mortality as death from any cause within 30 days of the date of admission for the index stroke hospitalization.

Denominator time window: This measure will use index admissions during a three-year period for the denominator.

S.6. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm.
The measure outcome is death from any cause within 30 days of the admission date of the index admission. As currently specified, we identify deaths for FFS Medicare patients, age 65 years and older in the Medicare Enrollment Database (EDB).

S.7. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The cohort includes inpatient admissions for Medicare FFS patients, age 65 years and older, who were discharged from non-federal, short-term, acute care hospitals with a principal discharge diagnosis of acute ischemic stroke.

Additional details are provided in S.9 Denominator Details.

S.8. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*
Populations at Risk, Senior Care

S.9. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, specific data collection items/responses , code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
The denominator includes all Medicare FFS beneficiaries, age 65 and over with a principal discharge diagnosis of acute ischemic stroke. To be included in the measure cohort used in public reporting, patients must meet the following additional inclusion criteria:

1. Enrolled in Medicare fee-for-service (FFS) during the index admission;
2. Not transferred from another acute care facility; and
3. Enrolled in Part A and Part B Medicare for the 12 months prior to the date of index admission.

ICD-9-CM codes that define the patient cohort:
433.01 Occlusion and stenosis of basilar artery with cerebral infarction
433.11 Occlusion and stenosis of carotid artery with cerebral infarction
433.21 Occlusion and stenosis of vertebral artery with cerebral infarction
433.31 Occlusion and stenosis of multiple and bilateral precerebral arteries with cerebral infarction
433.81 Occlusion and stenosis of other specified precerebral artery with cerebral infarction
433.91 Occlusion and stenosis of unspecified precerebral artery with cerebral infarction
434.01 Cerebral thrombosis with cerebral infarction
434.11 Cerebral embolism with cerebral infarction
434.91 Cerebral artery occlusion, unspecified with cerebral infarction
436 Acute, but ill-defined, cerebrovascular disease

ICD-10 codes that define the patient cohort:
I63.22 Cerebral infarction due to unspecified occlusion or stenosis of basilar arteries
I63.139 Cerebral infarction due to embolism of unspecified carotid artery

I63.239 Cerebral infarction due to unspecified occlusion or stenosis of unspecified carotid arteries
I63.019 Cerebral infarction due to thrombosis of unspecified vertebral artery
I63.119 Cerebral infarction due to embolism of unspecified vertebral artery
I63.219 Cerebral infarction due to unspecified occlusion or stenosis of unspecified vertebral arteries
I63.59 Cerebral infarction due to unspecified occlusion or stenosis of other cerebral artery
I63.20 Cerebral infarction due to unspecified occlusion or stenosis of unspecified precerebral arteries
I63.30 Cerebral infarction due to thrombosis of unspecified cerebral artery
I63.40 Cerebral infarction due to embolism of unspecified cerebral artery
I63.50 Cerebral infarction due to unspecified occlusion or stenosis of unspecified cerebral artery
I67.8 Other specified cerebrovascular diseases

An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

S.10. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
The measure excludes admissions for patients:
1. With inconsistent or unknown vital status or other unreliable data;
2. Enrolled in the Medicare hospice program at any time in the 12 months prior to the index admission, including the first day of the index admission; and
3. Discharged against medical advice (AMA).
For patients with more than one admission for stroke in a given year, only one index admission for that condition is randomly selected for inclusion in the cohort.

S.11. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
1. Inconsistent vital status or unreliable data: We do not include stays for patients where the age is greater than 115, where the gender is neither male nor female, where the admission date is after the date of death in the Medicare Enrollment Database, or where the date of death occurs before the date of discharge but the patient was discharged alive.

2. Hospice enrollment in the 12 months prior to or on the index admission is identified using hospice data and the Inpatient Standard Analytic File (SAF). These patients are likely continuing to seek comfort measures only; thus, mortality is not necessarily an adverse outcome or signal of poor quality care for these patients.

3. Discharges against medical advice (AMA) are identified using the discharge disposition indicator. After all exclusions are applied, the measure randomly selects one index admission per patient per year for inclusion in the cohort so that each episode of care is mutually independent with the same probability of the outcome. For each patient, the probability of death increases with each subsequent admission, and therefore, the episodes of care are not mutually independent. Similarly, for the three year combined data, when index admissions occur during the transition between measure reporting periods (June and July of each year) and both are randomly selected for inclusion in the measure, the measure includes only the June admission. The July admissions are excluded to avoid assigning a single death to two admissions.

S.12. **Stratification Details/Variables** *(All information required to stratify the measure results including the stratification variables, definitions, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b)*
N/A

S.13. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in S.12 and for statistical model in S.14-15)
Statistical risk model
If other:

S.14. **Identify the statistical risk model method and variables** *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development and testing should be addressed with measure testing under Scientific Acceptability)*

Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).

The measure employs a hierarchical logistic regression model to create a hospital-level 30-day RSMR. In brief, the approach simultaneously models data at the patient and hospital levels to account for the variance in patient outcomes within and between hospitals (Normand & Shahian, 2007). At the patient level, it models the log-odds of hospital mortality within 30 days of discharge using age, selected clinical covariates, and a hospital-specific intercept. At the hospital level, the approach models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of mortality at the hospital, after accounting for patient risk. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Candidate and Final Risk-adjustment Variables:
This measure uses risk variables from both claims data and clinical data from patients' medical records. Candidate variables from the claims data were patient-level risk-adjustors that were expected to be predictive of mortality, based on empirical analysis, prior literature, and clinical judgment, including age, sex, and indicators of comorbidity and disease severity. Only claims variables in the current publically reported claims-based stroke mortality measure were considered as claims-based candidate variables. For each patient, covariates are obtained from claims records extending 12 months prior to and including the index admission. For this measure and the measure currently implemented by CMS, these risk-adjustment variables are identified using both inpatient and outpatient Medicare FFS claims data. We use condition categories (CCs), which are clinically meaningful groupings of more than 15,000 ICD-9-CM diagnosis codes (Pope et al., 2000). A file that contains a list of the ICD-9-CM codes and their groupings into CCs is attached in data field S.2b (Data Dictionary or Code Table). The model adjusts for case-mix differences based on the clinical status of patients at the time of admission. In addition, only comorbidities that convey information about the patient at admission or in the 12 months prior, and not complications that arise during the course of the index hospitalization, are included in the risk adjustment. Hence, we do not risk adjust for CCs that may represent adverse events of care and that are only recorded in the index admission. This list included a variable indicating whether a patient was transferred into an inpatient admission from the emergency department (ED). The ED transfer indicator variable is also a risk-adjustment variable derived from claims that is used in the current publically reported claims-based stroke mortality measure. It was added to the risk-adjustment model for that measure in 2013 because these patients are believed to have a higher risk of mortality. Therefore, we included this variable in the list of candidate variables for the hybrid stroke mortality measure as well.

In addition to the claims-derived candidate variables, we include data elements derived from patients' medical records as candidate variables. For development of the hybrid measure, we used data elements found in the American Heart Association/American Stroke Association (AHA/ASA) Get With The Guidelines (GWTG)-Stroke Registry as a surrogate for data that can be feasibly extracted from current EHR systems. To ensure the measures could be implemented using EHR data rather than registry data, we sought to ensure that all candidate clinical risk variables could be feasibly extracted from the EHR. For details on feasibility criteria, see section 2b2.4.

We also included the NIH Stroke Scale score in our list of candidate variables, which could be obtained from either claims or EHR systems. We used NIH Stroke Scale score values collected in the GWTG-Stroke Registry in order to include an assessment of stroke severity in the list of candidate variables.

The final set of risk adjustment variables for the hybrid stroke mortality measure (which uses both claims and electronic clinical data for risk adjustment) is:

Demographics
Age (continuous, per 5 units)

Additional Risk Factors
Transfer from another ED
NIH Stroke Scale score (continuous, per 5 units)
Blood glucose/10 (mg/dl) (electronic clinical data element only)

Heart rate (electronic clinical data element only)
Diastolic blood pressure (electronic clinical data element only)

Comorbidities
Congestive heart failure (CC 80)
Congenital cardiac/circulatory defects (CC 87-88)
Specified heart arrhythmias (CC 92)
Precerebral arterial occlusion and transient cerebral ischemia (CC 97)
Cerebral atherosclerosis and aneurysm (CC 98)
Metastatic cancer and acute leukemia and other major cancers (CC 7-8)
Protein-calorie malnutrition (CC 21)
Other significant endocrine and metabolic disorders (CC 22-24)
Disorders of the vertebrae and spinal discs (CC 39)
Other gastrointestinal disorders (CC 36)
Other musculoskeletal and connective tissue disorders (CC 43)
Iron deficiency and other/unspecified anemia and blood disease (CC 47)
Dementia or other specified brain disorders (CC 49-50)
Pneumonia (CC 111-113)
Decubitus ulcer of skin (CC 148)

Please note that while there are 21 discrete data elements here, many analytic tables include 24 risk variables because quadratic terms were used for the electronic clinical data variables due to the non-linearity of their relationship with the mortality outcome.

References:
Krumholz HM, Brindis RG, Brush JE, et al. 2006. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. Circulation 113: 456-462.

Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.

Pope GC, et al. 2000. Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Care Financing Review 21(3): 93-118.

S.15. Detailed risk model specifications *(must be in attached data dictionary/code list Excel or csv file. Also indicate if available at measure-specific URL identified in S.1.)*
Note: Risk model details (including coefficients, equations, codes with descriptors, definitions), should be provided on a separate worksheet in the suggested format in the Excel or csv file with data dictionary/code lists at S.2b.
Available in attached Excel or csv file at S.2b

S.15a. Detailed risk model specifications (*if not provided in excel or csv file at S.2b*)

S.16. Type of score:
Rate/proportion
If other:

S.17. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.18. Calculation Algorithm/Measure Logic *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk*

adjustment; etc.)
The measure estimates hospital-level, 30-day, all-cause RSMRs following hospitalization for stroke using hierarchical logistic regression models. In brief, the approach simultaneously models data at the patient and hospital levels to account for variance in patient outcomes within and between hospitals (Normand and Shahian, 2007). At the patient level, it models the log-odds of mortality within 30 days of index admission using age, selected clinical covariates, and a hospital-specific intercept. At the hospital level, it models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept represents the underlying risk of a mortality at the hospital, after accounting for patient risk. The hospital-specific intercepts are given a distribution to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

The RSMR is calculated as the ratio of the number of "predicted" to the number of "expected" deaths at a given hospital, multiplied by the national observed mortality rate. For each hospital, the numerator of the ratio is the number of deaths within 30 days predicted on the basis of the hospital's performance with its observed case mix, and the denominator is the number of deaths expected based on the nation's performance with that hospital's case mix. This approach is analogous to a ratio of "observed" to "expected" used in other types of statistical analyses. It conceptually allows for a comparison of a particular hospital's performance given its case mix to an average hospital's performance with the same case mix. Thus, a lower ratio indicates lower-than-expected mortality rates or better quality, and a higher ratio indicates higher-than-expected mortality rates or worse quality.

The "predicted" number of deaths (the numerator) is calculated by using the coefficients estimated by regressing the risk factors and the hospital-specific intercept on the risk of mortality. The estimated hospital-specific intercept is added to the sum of the estimated regression coefficients multiplied by the patient characteristics. The results are transformed and summed over all patients attributed to a hospital to get a predicted value. The "expected" number of deaths (the denominator) is obtained in the same manner, but a common intercept using all hospitals in our sample is added in place of the hospital-specific intercept. The results are transformed and summed over all patients in the hospital to get an expected value. To assess hospital performance for each reporting period, we re-estimate the model coefficients using the years of data in that period.

This calculation transforms the ratio of predicted over expected into a rate that is compared to the national observed mortality rate. The hierarchical logistic regression models are described fully in the original methodology report (Grosso et al., 2011).

References:
Normand S-LT, Shahian DM. 2007. Statistical and Clinical Aspects of Hospital Outcomes Profiling. Stat Sci 22(2): 206-226.

S.19. **Calculation Algorithm/Measure Logic Diagram URL or Attachment** *(You also may provide a diagram of the Calculation Algorithm/Measure Logic described above at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A. This measure is not based on a sample or survey.

S.21. **Survey/Patient-reported data** *(If measure is based on a survey, provide instructions for conducting the survey and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A. This measure is not based on a sample or survey.

S.22. **Missing data** (specify how missing data are handled, e.g., imputation, delete case.)
Required for Composites and PRO-PMs.
Claims data
Missing values are rare among variables used from claims data in this measure. However, for measure development, the NIHSS was pulled from the AHA/ASA GWTG-Stroke Registry. Collection of the NIHSS is now recommended in the AHA/ASA guidelines as Class I for care of patients admitted with acute ischemic stroke. Analysis of data from the GWTG-Stroke Registry indicate that hospitals are collecting these data on an increasing proportion of patients admitted for acute ischemic stroke. Among the measure cohort, NIH

Stroke Scale scores were available in 82.39% of patients with an admission for ischemic stroke from July 1, 2013 and June 30, 2014. This proportion increased from 70.78% of admissions occurring between July 1, 2011 and June 30, 2012. It is CMS's intention to use claims data or EHR data for NIH Stroke Scale scores when this measure is implemented. New ICD-10 codes for NIHSS scores will be available for use in October 2016. Reliability of the NIHSS data will be reassessed prior to implementation and after hospitals begin using these codes in Medicare claims.

Electronic clinical data
For measure development, the critical clinical data elements included in the risk model that will be implemented using EHR data, were drawn from the AHA/ASA GWTG-Stroke registry dataset. When the measure is implemented these data elements will be derived from hospital EHRs. We have empirically tested the feasibility of each of these data elements and have shown them to be consistently captured for nearly all adults hospitalized for acute ischemic stroke and extractable from hospital EHRs. In the instances where these data elements were missing for patients in the registry, we use multiple imputation to generate a range of plausible values for all missing data and estimate values for missing data.

In multiple imputation, missing variable values are predicted using other patient variables available. The predicted values are substituted for the missing values, which results in a full data set (the imputed data set) without any missing variables. By repeating this process multiple times, we get multiple imputed data sets. We then conduct analyses on and obtain results for each imputed data set. The results based on multiple data sets are combined to produce the overall final results. Because we do not rely on one particular plausible version of the value, we have multiple versions of the plausible values. In general, imputed values are not intended to be "guesses" of what any particular missing value might be; instead, multiple imputation is used to preserve the important characteristics of the underlying data set and the inherent relationships among the variables in the data set. The multiple imputation represents a random sample of the missing values according to the association of the non-missing values of all the variables considered. The resulting inferences of multiple imputation are statistically valid and reflective of the uncertainty due to missing values [He & Belin, 2014; Carpenter & Kenward; Rubin, 1987].
Five copies of imputation datasets were produced for the analyses, and then the results based on these data separately were aggregated according to the standard statistical methods for presentation and for the measure score calculation. The approach to handling missing variables will be updated for implementation.

References:
He R, Belin T. Multiple imputation for high-dimensional mixed incomplete continuous and binary data. Stat. Med. 2014;33:2251–2262.
Carpenter J, Kenward M. Wiley: Multiple Imputation and its Application - [Internet]. [cited 2015 May 18];Available from: http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470740523.html
Rubin DB. Frontmatter [Internet]. In: Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, Inc.; 1987 [cited 2015 May 15]. p. i–xxix.Available from: http://onlinelibrary.wiley.com/doi/10.1002/9780470316696.fmatter/summary

S.23. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.24.
Administrative claims, Electronic Clinical Data : Electronic Health Record, Electronic Clinical Data : Laboratory, Electronic Clinical Data : Registry, Other

S.24. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument e.g. name of database, clinical registry, collection instrument, etc.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
For measure implementation the data sources will be:
1. Medicare Part A inpatient and Part B outpatient claims: This data source contains claims data for FFS inpatient and outpatient services including: Medicare inpatient hospital care, outpatient hospital services, skilled nursing facility care, some home health agency services, as well as inpatient and outpatient physician claims for the 12 months prior to an index admission.

2. Medicare Enrollment Database (EDB): This database contains Medicare beneficiary demographic, benefit/coverage, and vital status information. This data source was used to obtain information on several inclusion/exclusion indicators such as Medicare status on admission as well as vital status. These data have previously been shown to accurately reflect patient vital status (Fleming

et al., 1992).

3. Electronic clinical data: The measure will be implemented using electronic clinical data from hospitals' EHRs for risk adjustment. Electronic clinical data includes laboratory results and vital signs at the patient level for all patients included in the cohort.

Reference:
Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs hospitals. Medical Care. 1992; 30(5): 377-91. Data sources for the all-payer update

S.25. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.26. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Facility

S.27. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital/Acute Care Facility
If other:

S.28. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
N/A

2a. Reliability – See attached Measure Testing Submission Form
2b. Validity – See attached Measure Testing Submission Form
NQF_2877_Hybrid_Stroke_Mortality_NQF_Testing_Attachment_v1.1.docx

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Hybrid Hospital 30-day, all-cause, risk standardized mortality rate (RSMR) following an acute ischemic stroke hospitalization with risk adjustment for stroke severity

Date of Submission: Click here to enter a date

Type of Measure:

☐ Composite – ***STOP – use composite testing form***	☒ Outcome (*including PRO-PM*)
☐ Cost/resource	☐ Process
☐ Efficiency	☐ Structure

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient

frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):
- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR
- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☒ clinical database/registry	☒ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☒ eMeasure (HQMF) implemented in EHRs	☒ eMeasure (HQMF) implemented in EHRs
☒ other: Census Data/American Community Survey	☒ other: Census Data/American Community Survey

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

The dataset used for testing included Medicare Parts A and B claims, the Medicare Enrollment Database (EDB), and data from the American Heart Association/American Stroke Association (AHA/ASA)'s Get With The Guidelines (GWTG)-Stroke Registry. The registry data were used as a surrogate for data that will eventually come from electronic health records (EHRs), or, in the case of stroke severity, from claims once ICD-10 codes for stroke severity are available and consistently used by hospitals. Additionally, census as well as claims data were used to assess socioeconomic factors and race (dual eligible and African American race variables obtained through enrollment data; Agency for Healthcare Research and Quality (AHRQ) socioeconomic status (SES) index score obtained through census data). The dataset used varies by testing type; see Section 1.7 for details.

1.3. What are the dates of the data used in testing?

The dates used vary by testing type; see Section 1.7 for details.

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other:

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

For this measure, hospitals are the measured entities. The testing dataset included a subset of 1,520 non-federal, acute inpatient US hospitals (including territories) that participate in the American Heart Association/American Stroke Association (AHA/ASA)'s Get With The Guidelines (GWTG)-Stroke Registry. The number of measured entities varies slightly by the type of testing performed; see Section 1.7 for details.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

The number of admissions/patients varies by testing type; see Section 1.7 for details.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

The datasets, dates, number of measured hospitals, and number of admissions used in each type of testing are as follows:

For reliability testing

The reliability of the model was tested by randomly selecting 50% of **Dataset 1** (development dataset) and developing a risk-adjusted model for this group. We then developed a second model for the remaining 50% of patients (validation cohort) and compared the two. Thus, for reliability testing, we randomly split **Dataset 1** into two samples.

For measure development purposes only, we used two linked data sources to create **Dataset 1**: Medicare Administrative claims and the AHA/ASA's GWTG-Stroke Registry. Registry data were used to obtain the National Institutes of Health (NIH) Stroke Scale scores and clinical risk variables that could in the future be extracted from EHRs.

Dataset 1 (development dataset): Merged Medicare Part A Inpatient and Outpatient and Part B Outpatient claims with GWTG-Stroke Registry stroke severity and clinical data
Dates of Data: July 1, 2011 – June 30, 2014
Number of Admissions: 188,975
Patient Descriptive Characteristics: average age=79.47, %male=43.39
Number of Measured Hospitals: 1,511

First half of split sample (development sample):
- Number of Admissions: 94,466
- Number of Measured Hospitals: 1,473

Second half of split sample (validation sample):
- Number of Admissions: 94,509
- Number of Measured Hospitals: 1,462

For feasibility and validity testing (Section 2b2)

Dataset 1 was used for measure validity testing

Three additional datasets were used to assess the feasibility and validity of several critical data elements.

Dataset 3was used for data element feasibility testing

Data was provided from the administrative and EHR data warehouses of a large integrated health care delivery system that serves over 3.3 million members. All hospitals in this dataset used an integrated EHR system that runs **Epic software**.

• Number of admissions in dataset: 381,980

• Number of hospitals: 21

• Patient Descriptive Characteristics: mean age = 58 with a standard deviation of 21 years; %female = 62.6

Dataset 4 was used for data element feasibility and validity testing

Data were electronically extracted from one hospital that used **Cerner** as their clinical EHR.

• Number of hospitals//unique patient admissions used for feasibility testing: 3//18,380

• Number of hospitals//admissions used for validity testing: 1// 368

Dataset 5 used for data element validity testing

Data were electronically extracted from one hospital that used **GE Centricity** for their clinical EHR.

• Number of patients in dataset: 391

• Number of hospitals: 1

<u>For testing of measure exclusions (Section 2b3)</u>

Dataset 1 (prior to exclusions being applied)
Number of Eligible Admissions: 217,723
Number of Eligible Measured Entities: 1,511

<u>For testing of measure risk adjustment (Section 2b4)</u>

Dataset 1 development and validation samples

<u>For testing to identify meaningful differences in performance (Section 2b5)</u>

Dataset 1

<u>For testing of socioeconomic status (SES) factors and race in risk models (Section 2b4.4b)</u>

Dataset 1 and **Dataset 2**: The American Community Survey (2008-2012)

We examined disparities in performance according to the proportion of patients in each hospital who were of African-American race and the proportion who were dual eligible for both Medicare and Medicaid insurances. We also used the AHRQ SES index score derived from the American Community Survey (2008-2012) (**Dataset 2**) to study the association between performance measures and socioeconomic status.

Data Elements

• African-American race and dual eligible status (i.e., enrolled in both Medicare and Medicaid) patient-level data are obtained from CMS enrollment data (**Dataset 1**)
• Validated AHRQ SES index score is a composite of 7 different variables found in the census data (**Dataset 2**, the American Community Survey [2008-2012])

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Sociodemographic status incorporates socioeconomic variables as well as race into a more concise term. However, given the fact that socioeconomic risk factors are distinct from race and should be interpreted differently, we have decided to refer to "socioeconomic status" and "race" separately throughout this form.

We selected socioeconomic status (SES) and race variables to analyze after reviewing the literature and examining available national data sources. There is a large body of literature linking various SES factors and African-American race to worse health status and higher mortality over a lifetime (Adler and Newman, 2002; Mackenbach et al., 2000; Tonne et al., 2005; van Oeffelen et al., 2012). Income, education, and occupational level are the most commonly examined variables. Studies examining stroke mortality have suggested an association with race and SES factors (Khan et al., 2011; Pedigo et al., 2011; Glymour et al., 2009; Clark et al., 2011; Boan et al., 2014, Howard et al., 2011); however, the literature contains few studies directly examining how different SES factors or race might influence the likelihood of older, insured, Medicare patients of dying within 30 days of an admission for a stroke. The causal pathways for SES and race variable selection are described below in Section 2b4.3.

The SES and race variables used for analysis were:

- Dual eligible status (**Dataset 1**)
- African-American race (**Dataset 1**)
- AHRQ-validated SES index score using 5-digit zip code data (percentage of people in the labor force who are unemployed, percentage of people living below poverty level, median household income, median value of owner-occupied dwellings, percentage of people ≥25 years of age with less than a 12th-grade education, percentage of people ≥25 years of age completing ≥4 years of college, and percentage of households that average ≥1 people per room) (**Dataset 2**)

References:

Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. *Health affairs (Project Hope).* 2002; 21(2):60-76.

Boan AD, Feng WW, Ovbiagele B, et al. Persistent racial disparity in stroke hospitalization and economic impact in young adults in the buckle of stroke belt. Stroke; a journal of cerebral circulation. Jul 2014;45(7):1932-1938.

Clark CJ, Guo H, Lunos S, et al. Neighborhood cohesion is associated with reduced risk of stroke mortality. Stroke; a journal of cerebral circulation. May 2011;42(5):1212-1217.

Glymour MM, Kosheleva A, Boden-Albala B. Birth and adult residence in the Stroke Belt independently predict stroke mortality. Neurology. Dec 1 2009;73(22):1858-1865.

Howard VJ, Kleindorfer DO, Judd SE, et al. Disparities in stroke incidence contributing to disparities in stroke mortality. Ann Neurol 2011;69:619–627.

Khan JA, Casper M, Asimos AW, et al. Geographic and sociodemographic disparities in drive times to Joint Commission-certified primary stroke centers in North Carolina, South Carolina, and Georgia. Preventing chronic disease. Jul 2011;8(4):A79.

Mackenbach JP, Cavelaars AE, Kunst AE, Groenhof F. Socioeconomic inequalities in cardiovascular disease mortality; an international study. *European heart journal.* 2000; 21(14):1141-1151.

Pedigo A, Seaver W, Odoi A. Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: fuzzy cluster and discriminant analyses approaches. PloS one. 2011;6(7):e22693.

Tonne C, Schwartz J, Mittleman M, Melly S, Suh H, Goldberg R. Long-term survival after acute myocardial infarction is lower in more deprived neighborhoods. *Circulation.* Jun 14 2005; 111(23):3063-3070.

van Oeffelen AA, Agyemang C, Bots ML, et al. The relation between socioeconomic status and short-term mortality after acute myocardial infarction persists in the elderly: results from a nationwide study. *European journal of epidemiology.* Aug 2012; 27(8):605-613.

_____

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Data Element Reliability: Claims-derived data elements

In constructing the measure, we aim to utilize only those data elements from the claims that have both face validity and reliability. We generally avoid the use of fields that are thought to be coded inconsistently across hospitals or providers. Specifically, we use fields that are consequential for payment and which are audited. We identify such variables through empiric analyses and our understanding of CMS auditing and billing policies and seek to avoid variables which do not meet this standard. For example, "discharge disposition" is a variable in Medicare claims data that is not thought to be a reliable variable for identifying a transfer between two acute care facilities. Thus, we derive a variable using admission and discharge dates as a surrogate for "discharge disposition" to identify hospital admissions involving transfers. This allows us to identify these admissions using variables in the claims data which have greater reliability than the "discharge disposition" variable. In addition, CMS has in place several hospital auditing programs used to assess overall claims code accuracy, to ensure appropriate billing, and for overpayment recoupment. CMS routinely conducts data analysis to identify potential problem areas and detect fraud, and audits important data fields used in our measures, including diagnosis and procedure codes and other elements that are consequential to payment.

In addition to comorbidity data from claims, this measure includes the National Institutes of Health (NIH) Stroke Scale score to adjust for stroke severity, which will be included in ICD-10 CM codes beginning in October 2016. For development, we used NIH Stroke Scale (and several other clinical data elements) from the AHA/ASA's GWTG-Stroke Registry. However, it is CMS's intention to use claims data or EHR data, and not registry data, for NIH Stroke Scale scores and to use EHR data, not registry data for the other clinical data elements when this measure is implemented. New ICD-10 codes for NIH Stroke Scale scores will be available for use in October 2016. Given clinical guidelines and extensive prior testing and use of this score we expect reliable coding of NIH stroke severity scores, but CMS will monitor its use as a part of this measure implementation.

Finally, we assess the reliability of the claims data elements by comparing model variable frequencies and odds ratios from logistic regression models in the development and validation samples (July 1, 2011-June 30, 2014, **Dataset 1**).

Data Element Reliability: Electronic clinical data elements

See section 2b2 for validity testing of data elements (and see attached "Hospital-Level Measures of 30-Day Mortality Following Acute Ischemic Stroke Hospitalization that Incorporate Risk Adjustment for Stroke Severity Technical Report").

Measure Score Reliability

The reliability of a measurement is the degree to which repeated measurements of the same entity agree with each other. For measures of hospital performance, the measured entity is the hospital, and reliability is the extent to which repeated measurements of the same hospital give similar results. In line with this thinking, our approach to assessing reliability is to consider the extent to which assessments of a hospital using different but randomly selected subsets of patients produces similar measures of hospital performance. That is, we take a "test-retest" approach in which hospital performance is measured once using a random subset of patients, then measured again using a second random subset exclusive of the first, and finally comparing the agreement between the two resulting performance measures across hospitals (Rousson et al., 2002).

For test-retest reliability, we combined index admissions from successive measurement periods into one dataset, randomly sampled half of patients within each hospital, calculated the measure for each hospital, and repeated the calculation using the second half. Thus, each hospital is measured twice, but each measurement is made using an entirely distinct set of patients. To the extent that the calculated measures of these two subsets agree, we have evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement we calculated the intra-class correlation coefficient (ICC), and assessed the values according to conventional standards (Landis and Koch, 1977; Shrout and Fleiss, 1979). Specifically, we used **Dataset 1** split samples and calculated the RSMR for each hospital for each sample. The agreement of the two RSMRs was quantified for hospitals in each sample using the intra-class correlation as defined by ICC (2,1) by Shrout and Fleiss (1979).

Using two independent samples provides a stringent estimate of the measure's reliability, compared with using two random but potentially overlapping samples which would exaggerate the agreement. Moreover, because our final measure is derived using hierarchical logistic regression, and a known property of hierarchical logistic regression models is that smaller volume hospitals contribute less 'signal,' a split sample using a single measurement period would introduce extra noise. This leads to an underestimate in the actual test-retest reliability that would be achieved if the measure were reported using the full measurement period, as evidenced by the Spearman Brown prophecy formula (Spearman, 1910; Brown, 1910). We use this to estimate the reliability of the measure if the whole cohort were used, based on an estimate from half the cohort.

References:

Brown, W. (1910). Some experimental results in the correlation of mental abilities. British Journal of Psychology, 3, 296–322.

Landis J, Koch G, The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.

Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test–retest reliability of continuous measurements. Statistics in Medicine 2002;21:3431-3446.

Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979;86:420-428.

Spearman, Charles, C. (1910). Correlation calculated from faulty data. British Journal of Psychology, 3, 271–295.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Data element reliability results: Claims-derived data elements (**Dataset 1**)

The frequency of model variables remained relatively constant between 2011 and 2014, with no model variables increasing or decreasing by more than 2%.

Analysis of data from the Get With The Guidelines (GWTG)-Stroke Registry indicate that hospitals are collecting these data with increasing frequency for patients admitted for acute ischemic stroke. Among the measure cohort, NIH Stroke Scale scores were available in 82.39% of patients with an admission for ischemic stroke from July 1, 2013 and June 30, 2014. This proportion increased from 70.78% of admissions occurring between July 1, 2011 and June 30, 2012.

Examination of the odds ratios for each risk variable in the model shows that, overall, the odds ratios for individual risk variables remained relatively constant across three years (in attached Data Dictionary and Code Table).

For the model variable frequencies and risk variable odds ratios, see field S.2b (Data Dictionary or Code Table).

Measure Score Reliability Results (**Dataset 1**)

There were 188,975 admissions in the measure cohort, with 94,466 in one randomly selected sample (development sample) and 94,509 in the other sample (validation sample). The agreement between the two RSMRs for each hospital was 0.561, which according to the conventional interpretation is “moderate” (Landis & Koch, 1977).

Note that this analysis was limited to hospitals with 12 or more cases in each split sample. The intra-class correlation coefficient is based on a split sample of three years of data, resulting in a volume of patients in each sample equivalent to only 1.5 years of data, whereas the measure is intended to be reported with the full three years of data.

Reference:

Landis J, Koch G. The measurement of observer agreement for categorical data, Biometrics 1977;33:159-174.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The stability over time of the odds ratios and model variable frequencies suggests that the underlying data elements are reliable. The ICC score demonstrates moderate agreement across samples using a conservative approach to assessment for the measure score.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

- ☒ **Empirical validity testing**
- ☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Validity of EHR Data Elements

Several critical clinical data elements in the registry dataset were derived from patients' medical records and used for risk adjustment of the hybrid stroke mortality measure. When this measure is implemented, CMS intends to obtain these critical data elements from hospital EHRs, merge the data with claims, and to calculate and report measure results. We tested the validity of electronic extraction of these critical data elements as part of a more comprehensive evaluation of a larger set of core clinical data elements (CCDEs). The CCDE are a set of 21 EHR data elements that are captured on nearly all adults (plus Troponin, which is a condition-specific CCDE for patients with acute myocardial infarction) admitted to acute care hospitals, are easily extracted from EHRs, and can be used to risk adjust hospital outcome measures for a variety of conditions and procedures. All three of the critical data elements used in the hybrid stroke mortality measure are included in the CCDE. Testing of the CCDE involved three phases: 1) identification of potentially feasible clinical data through qualitative assessment, 2) empirical feasibility testing of several clinical data elements in a large multi-site database, and 3) further feasibility testing and validity testing of the CCDE at two hospital sites in two different health systems using two different EHR software systems.

Phase 1: Identification of potentially feasible clinical data through qualitative assessment

In order to identify the CCDEs for risk adjustment of hospital outcome measures for adult patients, we first conducted a qualitative assessment of the reliable capture, accuracy, and extractability of categories and subcategories of clinical data as defined by the Quality Data Model (QDM) (e.g., vital signs, laboratory test results). We established a set of criteria to assess the consistency of data capture, relevance to hospital quality measures, and extractability from health records.

Data Capture Criteria:

Obtained consistently under current practice. Routinely collected for patients admitted to the hospital under current clinical practice and EHR workflows.

Captured with a standard definition. Consistent conceptual understanding, method of collection, and units of measurement.

Entered in a structured field. Captured in numerical, pseudo-numerical, or list format.

Data Extraction Criteria:

Encoded consistently. Can be linked to a standard and uniform coding structure such as ICD-9 or Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT).

Extractable from the EHR. Can be readily and consistently identified and exported from current EHR databases.

Exported with metadata. Additional information such as time stamps and reference values that are needed for interpretation are consistently available.

These criteria are aligned with those established in the NQF's eMeasure Feasibility Assessment Report as well as the NQF feasibility criteria (see included feasibility score card). The NQF report emphasized four key aspects of feasibility. First, data should be structured or easily converted to a structured and interpretable format. Second, data should be accurate. Third, data should be easily associated with a standard set of codes to ensure consistent extraction across EHR environments. Finally, data should not require changes to current clinical practice or workflows.

We then convened a Technical Expert Panel (TEP) to apply these criteria to categories and subcategories of clinical data based on the Quality Data Model (QDM). We asked TEP members to consider only the context of adult hospitalized patients when making their assessments. Data categories and subcategories were rated on each feasibility criterion independently by TEP members. The ratings were tallied and TEP members met to discuss and resolve areas of disagreement. Through this process the TEP identified a list of data subcategories that were potentially feasible for use in hospital outcome measures. The CCDEs were derived from only those categories for which the TEP reached consensus agreement on feasibility.

Phase 2: Empirical feasibility testing using a large multi-site database

In **Dataset 3**, we next directly examined the feasibility of clinical data elements from the subcategories identified by the TEP as feasible (for all adult inpatient admissions). This dataset was derived from a 21 facility hospital system that uses **Epic** for their EHR. All clinical data were extracted from their Epic EHR system and clinical laboratory database. We examined all admissions in **Dataset 3** between 2010 and 2011. We analyzed clinical data elements to determine the format of the data, the consistency and timing of capture, and the accuracy of the data elements. We examined the data elements across conditions, hospitals, and point of hospital entry. We tested several data elements that met the feasibility criteria in models predicting 30-day mortality following admission for several common medical conditions. The complete list of 21 (plus Troponin) core clinical data elements were derived from these analyses.

To verify that the findings from our analysis of **Dataset 3** were generalizable to other hospitals and electronic health systems, we partnered with Premier Inc., a collaborative healthcare alliance of approximately 2,900 U.S. community hospitals focused on measuring and improving their members' quality outcomes and safely reducing healthcare costs. We administered a survey to four of their member hospital systems that used a variety of EHR systems to confirm the availability of the clinical data elements.

Phase 3: Further feasibility testing and validity testing at two hospital sites the CCDE (including critical data elements for the hybrid stroke mortality measure)

In Phase 3, we developed electronic specifications (e-specifications) using the Measure Authoring Tool (MAT), and analyzed extracted data from electronic health records (EHRs). We assessed the ability of hospitals to use the e-specifications to query and electronically extract CCDEs from the EHR, for all adult inpatient admissions occurring over the course of one year. Additional feasibility testing confirmed the numeric structure and assessed the rate of capture and timing of the clinical data elements in 18,380 hospital admissions to three hospitals in a single health system using a **Cerner** EHR (**Dataset 4**). Validity testing assessed the accuracy of the electronically extracted CCDEs compared to the same CCDEs gathered through manual medical record abstraction in a subset of 368 charts identified in the data query in **Dataset 4**, and 391 charts identified in the data query in **Dataset 5 (GE Centricity** EHR**)**.

Chart Abstraction: We calculated the number of admissions that needed to be randomly sampled from the EHR dataset and manually abstracted to yield a statistical margin of error (MOE) of 5% and a confidence level of 95% for the match rates between the two data sources. Sites then used an Access-based manual abstraction tool provided (along with training) to manually abstract the CCDEs from the random samples identified through the EHR data query. The manual chart abstraction data is considered the "gold standard" for the purpose of this analysis.

Validity Testing: We conducted validity testing on all 3 of the critical EHR data elements in the hybrid stroke mortality measure. For each continuous data element, we were only interested in the case where the electronic abstraction value exactly matched the manual abstraction value. We therefore only calculated the raw agreement rate between data from electronic and manual chart abstraction. For simple data values, we believe taking this approach, as compared to reporting a statistical tests of accuracy, better reflects the concept of matching exact data values rather than calculated measure results. Therefore, we do not report statistical testing of the accuracy of the EHR derived data value as compared with the abstracted value, as such tests do not apply to simple data values. Instead, we present below in Table 3 the exact match in the data value as well as the time and date stamp associated with that value. The 95% confidence level was established based on the size of the sample of manually abstracted charts and reflects the match rate using the criterion that both the data value and metadata match exactly.

Validation Against Other Risk Models and Registry Data

The hybrid model we developed is the model being proposed for measure endorsement in this application, the hybrid risk-adjustment model. This model uses a combination of claims data (demographics, comorbidities, and patient medical

history) and electronic clinical data (demographics, laboratory results, and vital signs), as well as the NIH Stroke Scale score (which can be captured using administrative claims or electronic clinical data).

We compared the hybrid risk model to two additional, but harmonized 30-day hospital-level ischemic stroke mortality measures. All three measures include the first-captured NIH Stroke Scale score in their respective risk models and use Medicare administrative claims data to derive the cohort and the Medicare Enrollment Database to assess the mortality outcome.

Measure validity was tested through comparison of this hybrid risk adjustment model with these other two other approaches to risk adjustment of the stroke mortality measure, and through use of established measure development guidelines. We compared model performance to that of:

1. An updated claims-only risk-adjustment model that includes only variables obtained from administrative claims data (demographics, comorbidities, and patient medical history) and the NIH Stroke Scale score (which can be captured using administrative claims data starting in October 2016).

2. A clinical-only risk-adjustment model that includes only variables from electronic clinical data (demographics, laboratory results, and vital signs), as well as the NIH Stroke Scale score (which can be captured using electronic clinical data).

For the derivation of all three models, we linked cases from the registry to the corresponding administrative data from Medicare claims and mortality data from the Medicare enrollment database (**Dataset 1**). The GWTG-Stroke Registry draws data from medical records and has been shown to be reliable through the studies comparing registry data to chart abstraction (Xian et al., 2012). Each of the three cohorts for the three risk models used the same inclusion/exclusion criteria and a risk-adjustment (statistical modeling) strategy and only differed with respect to the risk variables used. Because only patients aged 65 years and older were included, and some data were excluded based on linkage and other factors, a total of 188,975 stroke hospitalizations were used in the analysis. We compared the model discrimination and the correlation in hospital performance results for each model compared with the hybrid stroke mortality model.

<u>Validity Indicated by Established Measure Development Guidelines:</u>

We developed this measure in consultation with national guidelines for publicly reported outcomes measures, with outside experts, and with the public. The measure is consistent with the technical approach to outcomes measurement set forth in NQF guidance for outcomes measures (National Quality Forum, 2010), CMS Measure Management System (MMS) guidance, and the guidance articulated in the American Heart Association scientific statement, “Standards for Statistical Models Used for Public Reporting of Health Outcomes” (Krumholz, Brindis, et al., 2006).

<u>Validity as Assessed by External Groups:</u>

Throughout measure development, we obtained expert and stakeholder input via regular discussions with an advisory working group and a 30-day public comment period in order to increase transparency and to gain broader input into the measure.

The working group was assembled, and regular meetings were held throughout the development phase. The working group was tailored for development of this measure and consisted of clinicians (neurologists and cardiologists) and other professionals with expertise in biostatistics, measure methodology, and quality improvement. The working group meetings addressed key issues related to measure development, including the deliberation and finalization of key decisions (e.g., defining the measure cohort and outcome) to ensure the measure is meaningful, useful, and well-designed. The working group provided a forum for focused expert review and discussion of technical issues during measure development.

Following completion of the preliminary model, we solicited public comment on the measure through the CMS site link: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/CallforPublicComment.html. The public comments were then posted publicly for 30 days. The resulting input was taken into consideration during the final stages of measure development and contributed to minor modifications to the measure.

Citations:

Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. *PLoS One* 2011;6(4):e17401.

Keenan PS, Normand SL, Lin Z, et al. An administrative claims measure suitable for profiling hospital performance on the basis of 30-day all-cause readmission rates among patients with heart failure. *Circulation* 2008;1(1):29-37.

Krumholz HM, Brindis RG,Brush JE, et al. Standards for Statistical Models Used for Public Reporting of Health Outcomes: An American Heart Association Scientific Statement From the Quality of Care and Outcomes Research Interdisciplinary Writing Group: Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council Endorsed by the American College of Cardiology Foundation. *Circulation.* January 24, 2006 2006;113(3):456-462.

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation* 2006;113(13):1683-92.

Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with heart failure. *Circulation* 2006;113:1693-1701.

National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report http://www.qualityforum.org/projects/Patient_Outcome_Measures_Phases1-2.aspx. Accessed August 19, 2010.

Shahian DM, He X, O'Brien S, et al. Development of a Clinical Registry-Based 30-Day Readmission Measure for Coronary Artery Bypass Grafting Surgery. Circulation 2014; DOI: 0.1161/CIRCULATIONAHA.113.007541. Published online before print June 10, 2014

Suter L, Wang C, Araas M, et al. Hospital-Level 30-Day All-Cause Unplanned Readmission Following Coronary Artery Bypass Graft Surgery (CABG): Updated Measure Methodology Report. 2014; http://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890352615&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DRdmsn_CABG_MeasMethd_Rpt_060314.pdf&blobcol=urldata&blobtable=MungoBlobs. Accessed November 4, 2015.

ICD-9 to ICD-10 Conversion

Statement of Intent

[X] Goal was to convert this measure to a new code set, fully consistent with the intent of the original measure.

[ ] Goal was to take advantage of the more specific code set to form a new version of the measure, but fully consistent with the original intent.

[ ] The intent of the measure has changed.

Process of Conversion

ICD-10 codes were identified using 2015 GEM mapping software. We enlisted the help of clinicians with expertise in relevant areas to select and evaluate which ICD-10 codes map to the ICD-9 codes currently in use for this measure. An ICD-9 to ICD-10 crosswalk is attached in field S.2b. (Data Dictionary or Code Table).

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Validity of EHR Data Elements

Phase 1: TEP Survey Results

The TEP identified seven subcategories (or datatypes) from the Quality Data Model (QDM) of EHR data that they considered feasible for adult hospitalized patients. They were Encounter Performed, Patient Characteristics including birth date and sex, Physical Examination Findings for vital signs only, Diagnostic Study Order, Diagnostic Study Performed, Medication Discharge, and Laboratory Test Result. We limited the CCDE to data elements to only 4 categories Encounter Performed, Patient Characteristics, Physical Examination Findings for vital signs only, and Laboratory Test Results, which are unlikely to be reflective of care quality and therefore are thought to be both feasible to extract and appropriate for risk adjustment.

Phase 2: Feasibility Testing Results

Dataset 3:The consistency of data capture of the critical data elements included in the hybrid stroke mortality measure for all adult hospitalized patients in **Dataset 3** are presented in Table 1 and Table 2. These tables show consistent capture of all three data elements.

In addition, the four Premier member hospitals reported all reported that the CCDE were: captured in the inpatient EHR; captured in the emergency department EHR; recorded in a structured format; extracted for reporting; extracted for other purposes; and time and date stamps captured.

Table 1 Proportion of Episodes with Captured Vital Signs at Various Time points (Dataset 3, Epic EHR)

Vital Sign Finding – Full Cohort	Total with Finding and Timestamp %	Within 2 Hours %	Within 6 Hours %	Within 12 Hours %
Basic vital signs				
Heart rate	99.7	96.8	99.4	99.6
Diastolic blood pressure	99.7	96.7	99.3	99.6

Table 2 Proportion of Admissions with Laboratory Results at Various Time Points (Dataset 3, Epic EHR)

Lab Test Result – Full Cohort	Total with Result and Timestamp (%)	Within 2 Hours (%)	Within 6 Hours (%)	Within 12 Hours (%)	Within 24 Hours (%)
Glucose	72.0	49.7	57.6	60.6	70.0

Phase 3: Further Feasibility and Validity Testing Results

Chart abstraction for validity testing was done in **Dataset 4 and Dataset 5.** Table 3 demonstrates the comparison between electronic and manual abstraction of data.

Table 3a: Proportion of Admissions with Data Element s Captured within 2 Hours for Vital Signs and 24 Hours for Lab Values (Dataset 4, Cerner EHR with 18,380 admissions)

Data Element/ CCDE	% Captured Hospital 1	% Captured Hospital 2	% Captured Hospital 3
Heart Rate (BPM)	94.27	80.54	84.48
Diastolic Blood Pressure (mmHG)	94.29	80.81	83.99
Glucose	77.64	80.93	81.32

Table 3b Proportion with Perfect match in Data Element Value Comparing EHR Extracted and Manual Medical Record Abstracted Data (Datasets 4 & 5; Cerner and GE Centricity EHRs)

Data Element/ CCDE	% Agreement Between Datasets (Number Matching/ Total Records With A Data Value)	95% Confidence Internal for Agreement	% Present in Electronic Extraction, Missing in Manual Abstraction (N)	% Present in Manual Abstraction, Missing in Electronic Extraction (N)	% Missing in Both Electronic Extraction and Manual Abstraction (N)
Dataset 4 (n=368)					
Heart rate (BPM)	95.55 (322/337)	92.76 - 97.49	0 (0.00)	8.42 (31)	0 (0.00)
Diast Blood Pressure (mmHG)	94.38 (319/338)	91.36 - 96.58	0 (0.00)	8.15 (30)	0 (0.00)
Glucose (mg/dL)	96.14 (274/285)	93.20 - 98.06	0 (0.00)	5.43 (20)	17.12 (63)
Dataset 5 (n=391)					
Heart rate (BPM)	57.45 (135/235)	50.85 - 63.85	0 (0.00)	39.39 (154)	0.51 (2)
Diast Blood Pressure (mmHG)	60.09 (137/228)	53.41 - 66.50	0.26 (1)	40.92 (160)	0.51 (2)
Glucose (mg/dL)	95.12 (78/82)	87.98 - 98.66	0 (0.00)	63.17 (247)	15.86 (62)

A post-validation review of the code used by the hospital in **Dataset 5**, revealed that the hospital experienced a number of errors. The most significant of which was extracting data only within an incorrect two-hour window for laboratory test results (the correct window was 24 hours). Additionally, physical exam (vital signs) data were extracted based on the date/time that results were documented rather than the date/time the physical exams were performed, driving down the accuracy of these data. However, post-validation review of the code used by the hospital in **Dataset 4** showed no such errors in the query executed. As a result the match rate was much higher.

Validation Against Other Risk Models and Registry Data

The performance of the claims-only risk-adjustment model and the clinical-only risk-adjustment model are similar to the performance of the hybrid risk-adjustment model (**Dataset 1,** development sample). The areas under the receiver operating characteristic (ROC) curve are 0.81 and 0.79, respectively, for the two models compared with 0.82 for the hybrid risk-adjustment model.

We estimated hospital-level RSMRs using the corresponding hierarchical logistic regression for each of the three models in the linked patient sample. We then examined the linear relationship between the estimates using regression techniques and weighting by the total number of cases in each hospital. The Pearson correlation coefficient of the standardized rates from the claims-only risk-adjustment model and the hybrid risk-adjustment model is 0.98631. The Pearson correlation coefficient of the standardized rates from the clinical-only risk-adjustment model and the hybrid risk-adjustment model is 0.96361.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Validity of EHR Data Elements

Feasibility Testing (Phases 1-3)

The critical data elements were demonstrated to be feasible through consensus of the TEP and direct examination of EHR data in 2 EHR systems (Epic and Cerner) and 24 total hospitals, establishing consistent capture of the CCDEs among adult hospitalized patients. In addition, we established the validity of electronic extraction of the CCDEs demonstrated by the high match rate when comparing EHR extracted and manual medical record abstracted CCDE values.

Measure Validity (**Dataset 1**)

The results between the hybrid stroke mortality model and the other two approaches to risk adjustment suggest that each approach shows similarly high discrimination. The ROC results were nearly identical and in line with other mortality models. In addition, the high correlation among the RSMRs calculated from all models shows that each model provides a similar or consistent measure result for hospitals.

Validity as Assessed by External Groups

The face validity testing results demonstrated working group and public comment agreement with overall face validity of the measure as specified.

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

All exclusions were determined by careful clinical review and have been made based on clinically relevant decisions and to ensure accurate calculation of the measure. To ascertain impact of exclusions on the cohort, we examined overall frequencies and proportions of the total cohort excluded for each exclusion criterion (**Dataset 1**). These exclusions are consistent with similar NQF-endorsed outcome measures. Rationales for the exclusions are detailed in data field S.10 (Denominator Exclusions).

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

In **Dataset 1** (prior to exclusions being applied):

Exclusion	N	%	Distribution across hospitals (N=1,511): Min, 25th, 50th, 75th percentile, max
1. Inconsistent or unknown vital status or other unreliable data	0	0.0%	(0, 0, 0, 0, 0)

2. Enrolled in the Medicare hospice program at any time in the 12 months prior to the index admission, including the first day of the index admission	1,527	0.78%	(0, 0, 0, 1, 15)
3. Discharged against medical advice (AMA)	439	0.22%	(0, 0, 0, 0, 6)

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusions 1 and 2 are necessary for valid calculation of the measure. **Exclusion 1** (patients with inconsistent or unknown vital status or other unreliable demographic [age and gender] data) accounts for 0.0% of all index admissions excluded from the initial index cohort. We do not include stays for patients where the age is greater than 115, where the gender is neither male nor female, where the admission date is after the date of death in the Medicare Enrollment Database, or where the date of death occurs before the date of discharge but the patient was discharged alive. **Exclusion 2** (patients enrolled in the Medicare hospice program any time in the 12 months prior to the index admission, including the first day of the index admission) accounts for 0.78% of all index admissions excluded from the initial index cohort. These patients are likely continuing to seek comfort measures only; mortality is not necessarily an adverse outcome or signal of poor quality care.

Exclusion 3 (patients who are discharged AMA) accounts for 0.22% of all index admissions excluded from the initial index cohort. This exclusion is needed for acceptability of the measure to hospitals, who do not have the opportunity to deliver full care and prepare the patient for discharge.

After all exclusions are applied, the measure randomly selects one index admission per patient per year for inclusion in the cohort so that each episode of care is mutually independent with the same probability of the outcome. For each patient, the probability of death increases with each subsequent admission, and therefore, the episodes of care are not mutually independent. Similarly, for the three year combined data, when index admissions occur during the transition between measure reporting periods (June and July of each year) and both are randomly selected for inclusion in the measure, the measure includes only the first admission. The subsequent admission is excluded to avoid assigning a single death to two admissions.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with** 21 **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.2. If an outcome or resource use measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Our approach to risk adjustment was tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).

The measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital-level 30-day RSMR. This approach to modeling appropriately accounts for the structure of the data (patients clustered within hospitals), the underlying risk due to patients' comorbidities, and sample size at a given hospital when estimating hospital mortality rates. In brief, the approach simultaneously models two levels (patient and hospital) to account for the variance in patient outcomes within and between hospitals (Normand and Shahian et al., 2007). At the patient level, each model adjusts the log-odds of mortality within 30-days of admission for age, selected clinical covariates and a hospital-specific intercept. The second level models the hospital-specific intercepts as arising from a normal distribution. The hospital intercept, or hospital-specific effect, represents the hospital contribution to the risk of mortality, after accounting for patient risk and sample size, and can be inferred as a measure of quality. The hospital-specific intercepts are given a distribution in order to account for the clustering (non-independence) of patients within the same hospital. If there were no differences among hospitals, then after adjusting for patient risk, the hospital intercepts should be identical across all hospitals.

Clinical Factors

We sought to develop a model that included key variables that were clinically relevant and based on strong relationships with the mortality outcome. To select candidate variables for the hybrid risk model, we began with the list of 42 administrative claims-based risk-adjustment variables included in the currently publicly reported Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure. These candidate variables were derived from: the index admission, with comorbidities identified from the index admission secondary diagnoses (excluding potential complications); 12-month pre-index inpatient data (for any condition); outpatient hospital data; Part B and physician data. In identifying these 42 variables for the current publically reported stroke mortality measure, we sought to develop a model that was parsimonious, using a grouper that is in the public domain for the 15,000+ ICD-9-CM codes we started with the 189 diagnostic groups included in the Hierarchical Condition Category (HCC) clinical classification system (Pope et al., 2000). The HCC clinical classification system was developed for CMS in preparation for all-encounter risk adjustment for Medicare Advantage (managed care) plans and represented a refinement of an earlier risk-adjustment method based solely on principal inpatient diagnosis. The HCC model makes use of all physician and hospital encounter diagnoses and was designed to predict a beneficiary's expenditures based on the total clinical profile represented by all of his/her assigned HCCs. Under the HCC algorithm, the 15,000+ ICD-9-CM diagnosis codes are first assigned to one of 804 mutually exclusive groupings ("DxGroups") and then subsequently aggregated into 189 condition categories (CCs). During development, we used the April 2008 version of the ICD-9-CM to CC assignment map, which is maintained annually by CMS and posted at www.qualitynet.org. We do not use the hierarchy and therefore refer to the CCs rather than HCCs.

To this initial list of 42 variables derived from claims, we added the NIH Stroke Scale and the list of 14 clinical data elements in the GWTG-Stroke Registry deemed feasibly extractable from the EHR. Our set of candidate variables from the claims included 2 demographic variables (age and gender), 39 CC-based variables, and an indicator variable for when a patient transferred into the hospital from the emergency department.

Our set of candidate variables from the GWTG-Stroke Registry, which were selected by a team of clinicians and analysts primarily based on their clinical relevance but with knowledge of their strength of association with the mortality outcome, included the NIH Stroke Scale, 6 laboratory results variables and 5 vital sign results.

To inform variable selection, a modified approach to stepwise logistic regression was performed. The developmental dataset was used to create 1,000 bootstrap samples. For each sample, we ran a logistic stepwise regression, with both backward and forward selection, that included the 56 candidate variables. The results were summarized to show the percentage of times that each of the candidate variables was significantly associated with mortality (at the $p<0.05$ level) in each of the 1,000 repeated samples (for example, a percentage of 90 would mean that the candidate variable was selected as significant at $p<0.05$ in 90% of the estimations). Variable selection rate for all of the variables was calculated for each copy of the multiple-imputed data, and variables were included in the final model if the minimum variable selection rate among the 5 copies of multiple imputed datasets was 90% or more. We also assessed the direction and magnitude of the regression coefficients. Based on visual inspection of the continuous candidate electronic clinical variable distributions, we included both the linear and quadratic terms of the variable when appropriate for bootstrapping.

The clinician team reviewed these results and decided to retain all risk-adjustment variables above a 90% cutoff, since they demonstrated a relatively strong association with mortality and were clinically relevant. If the quadratic version of the clinical variable reached the 90% selected rate, we included both the linear and quadratic versions in the model. This resulted in a claims and electronic clinical data risk-adjustment model that included 21 discrete risk variables (see Section 2b4.4a table of candidate variables).

Socioeconomic Factors and Race

We selected variables representing socioeconomic (SES) factors and race for examination based on a review of literature, conceptual pathways, and feasibility. In Section 1.8, we describe the variables that we considered and analyzed based on this review. Below we describe the pathways by which SDS and race may influence 30-day mortality.

Our conceptualization of the pathways by which patient SES or race affects 30-day mortality is informed by the literature.

Literature Review of Socioeconomic (SES) and Race Variables and Stroke Mortality

To examine the relationship between SES and race variables and hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following stroke hospitalization, a literature search was performed with the following exclusion criteria: articles published more than 10 years ago, international studies, articles using Veterans Affairs databases as the primary data source, and articles not explicitly focused on SES or race and stroke mortality. After applying the first exclusion criterion, 86 articles were reviewed by title and abstract, and 77 were excluded from full-text review. Nine articles were selected for full-text review. Among articles reviewed, several found an increased risk of in-hospital mortality associated with SES and/or race variables (Ovbiagele et al., 2010; Hasan et al., 2010), and others found an overall increase in mortality risk not confined to the inpatient hospitalization (Khan et al., 2011; Pedigo et al., 2011; Glymour et al., 2009; Clark et al., 2011; Boan et al., 2014, Howard et al., 2011). Some results showed no increased mortality risk associated with race/ethnicity but a risk conferred by SES (Hanchate et al., 2013).

Causal Pathways for Socioeconomic (SES) and Race Variable Selection

Although some recent literature evaluates the relationship between patient SES or race and the mortality outcome, few studies directly address causal pathways or examine the role of the hospital in these pathways. Moreover, the current literature examines a wide range of conditions and risk variables with no clear consensus on which risk factors demonstrate the strongest relationship with mortality. The SES factors that have been examined in the mortality literature can be categorized into three domains: (1) patient-level variables, (2) neighborhood/community-level variables, and (3) hospital-level variables. Patient-level variables describe characteristics of individual patients, and range from the self-reported or documented race or ethnicity of the patient to the patient's income or education level (Alter et al., 2014; Taksler et al., 2012). Neighborhood/community-level variables use information from sources such as the American Community Survey (ACS) as either a proxy for individual patient-level data or to measure environmental factors. Studies using these variables use one dimensional measures such as median household income or composite measures such as the Agency for Healthcare Research and Quality (AHRQ)-validated SES index score (Blum et al., 2014). Hospital-level variables measure attributes of the hospital which may be related to patient risk. Examples of hospital-level variables used in studies are ZIP code characteristics aggregated to the hospital level or the proportion of Medicaid patients served in the hospital.

The conceptual relationship, or potential causal pathways by which these possible SES risk factors influence the risk of mortality following an acute illness or major surgery, like the factors themselves, are varied and complex. There are at least four potential pathways that are important to consider.

1. **Relationship of socioeconomic (SES) factors or race to health at admission**. Patients who have lower income/education/literacy or unstable housing may have a worse general health status and may present for their hospitalization or procedure with a greater severity of underlying illness. These SES risk factors, which are characterized by patient-level or neighborhood/community-level (as proxy for patient-level) variables, may contribute to worse health status at admission due to competing priorities (restrictions based on job, lack of childcare), lack of access to care (geographic, cultural, or financial), or lack of health insurance. Given that these risk factors all lead to worse general health status, this causal pathway should be largely accounted for by current clinical risk-adjustment**.**

In addition to SES risk factors, studies have shown that worse health status is more prevalent among African-American patients compared with white patients. The association between race and worse health is in part mediated by the association between race and SES risk factors such as poverty or disparate access to care associated with poverty or neighborhood. The association between race and worse health status that is independent of poverty is mediated through bias in healthcare as well as in other facets of society.

2. **Use of low-quality hospitals**. Patients of lower income, lower education, or unstable housing have been shown not to have equitable access to high quality facilities because such facilities are less likely to be found in geographic areas with large populations of poor patients; thus patients with low income are more likely to be seen in lower quality hospitals, which can contribute to increased risk of mortality following hospitalization (Jha et al., 2011; Reames et al., 2014). Similarly African-American patients have been shown to have less access to high quality facilities compared with white patients (Skinner et al., 2005).

3. **Differential care within a hospital**. The third major pathway by which SES factors or race may contribute to mortality risk is that patients may not receive equivalent care within a facility. For example, African-American patients have been shown to experience differential, lower quality, or discriminatory care within a given facility (Trivedi et al., 2014). Alternatively, patients with SES risk factors such as lower education may require differentiated care – e.g. provision of lower literacy information – that they do not receive.

4. **Influence of SES on mortality risk outside of hospital quality and health status**. Some SES risk factors, such as income or wealth, may affect the likelihood of mortality without directly affecting health status at admission or the quality of care received during the hospital stay. For instance, while a hospital may make appropriate care decisions and provide tailored care and education, a lower-income patient may have a worse outcome post-discharge due to competing economic priorities or a lack of access to care outside of the hospital.

These proposed pathways are too complex to distinguish analytically. They also have different implications on the decision to risk adjust or not. We, therefore, first assessed if there was evidence of a meaningful effect on the risk model to warrant efforts to distinguish among these pathways. Based on this model and the considerations outlined in Section 1.8, the following SES and race variables were considered:

- African American race (as compared to all others)
- Dual eligible status
- AHRQ SES index score

We assessed the relationship between the SES variables and race with the outcome and examined the incremental effect in a multivariable model. For this measure, we also examined the extent to which the addition of any one of these variables improved model performance or changed hospital results. Given no meaningful improvement in the risk-model or change in performance scores we did not further seek to distinguish the causal pathways for these measures.

References:

Alter DA, Franklin B, Ko DT, et al. Socioeconomic status, functional recovery, and long-term mortality among patients surviving acute myocardial infarction. PloS one 2014; 8:e65130.

Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. Circulation Cardiovascular quality and outcomes 2014; 7:391-7.

Boan AD, Feng WW, Ovbiagele B, et al. Persistent racial disparity in stroke hospitalization and economic impact in young adults in the buckle of stroke belt. Stroke; a journal of cerebral circulation. Jul 2014;45(7):1932-1938.

Clark CJ, Guo H, Lunos S, et al. Neighborhood cohesion is associated with reduced risk of stroke mortality. Stroke; a journal of cerebral circulation. May 2011;42(5):1212-1217.

Glymour MM, Kosheleva A, Boden-Albala B. Birth and adult residence in the Stroke Belt independently predict stroke mortality. Neurology. Dec 1 2009;73(22):1858-1865.

Hanchate AD, Schwamm LH, Huang W, Hylek EM. Comparison of ischemic stroke outcomes and patient and hospital characteristics by race/ethnicity and socioeconomic status. Stroke; a journal of cerebral circulation. Feb 2013;44(2):469-476.

Hasan O, Orav EJ, Hicks LS. Insurance status and hospital care for myocardial infarction, stroke, and pneumonia. Journal of hospital medicine. Oct 2010;5(8):452-459.

Howard VJ, Kleindorfer DO, Judd SE, et al. Disparities in stroke incidence contributing to disparities in stroke mortality. Ann Neurol 2011;69:619–627.

Jha AK, Orav EJ, Epstein AM. Low-quality, high-cost hospitals, mainly in South, care for sharply higher shares of elderly black, Hispanic, and medicaid patients. Health affairs 2011; 30:1904-11.

Khan JA, Casper M, Asimos AW, et al. Geographic and sociodemographic disparities in drive times to Joint Commission-certified primary stroke centers in North Carolina, South Carolina, and Georgia. Preventing chronic disease. Jul 2011;8(4):A79.

Krumholz H, Normand S, Galusha D, et al. Risk-Adjustment Methodology for Hospital Monitoring/Surveillance and Public Reporting Supplement #1: 30-Day Mortality Model for Pneumonia. 2006. Available at: https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830. Accessed November 19, 2015.

Normand S-LT, Shahian DM. Statistical and Clinical Aspects of Hospital Outcomes Profiling. 2007/05 2007:206-226.

Ovbiagele B. Nationwide trends in in-hospital mortality among patients with stroke. Stroke; a journal of cerebral circulation. Aug 2010;41(8):1748-1754.

Pedigo A, Seaver W, Odoi A. Identifying unique neighborhood characteristics to guide health planning for stroke and heart attack: fuzzy cluster and discriminant analyses approaches. PloS one. 2011;6(7):e22693.

Reames BN, Birkmeyer NJ, Dimick JB, Ghaferi AA. Socioeconomic disparities in mortality after cancer surgery: failure to rescue. JAMA surgery 2014; 149:475-81.

Skinner J, Chandra A, Staiger D, Lee J, McClellan M. Mortality after acute myocardial infarction in hospitals that disproportionately treat black patients. Circulation 2005; 112:2634-41.

Taksler GB, Keating NL, Cutler DM. Explaining racial differences in prostate cancer mortality. Cancer 2012; 118:4280-9.

Trivedi AN, Nsa W, Hausmann LR, et al. Quality and equity of care in U.S. hospitals. The New England journal of medicine 2014; 371:2298-308.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Below are tables showing the candidate variables and the final variables that were included more than 90% of the time for all copies of the imputed data and therefore retained in the final model.

Claims-Derived Candidate Variables

Variable	Code(s)
NIH Stroke Scale score (continuous)	n/a
Age minus 65 (years above 65, continuous)	n/a
Male	n/a
Transfer from another ED	n/a
Congestive heart failure	CC 80
Valvular or rheumatic heart disease	CC 86
Congenital cardiac/circulatory defects	CC 87-88

Variable	Code(s)
Hypertensive heart disease	CC 90
Specified arrhythmias	CC 92
Cerebral hemorrhage	CC 95
Ischemic or unspecified stroke	CC 96
Precerebral arterial occlusion and transient cerebral ischemia	CC 97
Cerebral atherosclerosis and aneurysm	CC 98
Hemiplegia/hemiparesis	CC 100
History of infection	CC 1, 3-6
Metastatic cancer, acute leukemia and other severe cancers	CC 7-8
Lymphatic, head and neck, brain, and other major cancers; breast, colorectal and other major cancers	CC 9-13
Protein-calorie malnutrition	CC 21
Other significant endocrine and metabolic disorders	CC 22-24
Other gastrointestinal disorders	CC 36
Disorders of the vertebrae and spinal discs	CC 39
Osteoarthritis of hip or knee	CC 40
Other musculoskeletal and connective tissue disorders	CC 43
Iron deficiency or other unspecified anemias and blood disease	CC 47
Dementia or other specified brain disorders	CC 49-50
Major psychiatric disorders	CC 54-56
Quadriplegia, other extensive paralysis	CC 67-69
Multiple sclerosis	CC 72, 76
Seizure disorders and convulsions	CC 74
Hypertension	CC 89, 91
Vascular disease and complications	CC 104-105
Chronic obstructive pulmonary disease (COPD)	CC 108
Pneumonia	CC 111-113
Pleural effusion/pneumothorax	CC 114
Other eye disorders	CC 124
Other ear, nose, throat, and mouth disorders	CC 127
Dialysis status	CC 130
Renal failure	CC 131
Urinary tract infection	CC 135
Male genital disorders	CC 140
Decubitus ulcer of skin	CC 148
Chronic ulcer of skin, except decubitus	CC 149
Other dermatological disorders	CC 153

Registry Data Candidate Variables

Description
Age minus 65 (years above 65, continuous)

Male
NIH Stroke Scale (continuous)
Transfer from ED
Total cholesterol
HDL
LDL
Triglycerides
Blood glucose
INR
Serum creatinine
Heart rate
SBP
DBP
Weight
Mode of hospital arrival

Final Model Variables: Hybrid Model with Claims and Registry Data for Risk Adjustment (variables meeting criteria in field 2b4.3)

Variable	Code(s)	07/2011 – 06/2014 Coefficients (SE)	07/2011 – 06/2014 OR (95% CI)
Age (continuous, per 5 units)	--		1.35 (1.33, 1.36)
Transfer from another ED	--		1.44 (1.34, 1.54)
NIH Stroke Scale score (continuous, per 5 units)	--		1.56 (1.54, 1.58)
Blood glucose/10 (mg/dl): Linear	--	0.0481 (0.0065)	--
Blood glucose/10 (mg/dl): Square*	--	-0.0007 (0.0002)	--
Heart rate: Linear	--	-0.0024 (0.0054)	--
Heart rate: Square*	--	0.0001 (0.0000)	--
DBP: Linear	--	-0.0281 (0.0045)	--
DBP: Square*	--	0.0002 (0.0000)	--
Congestive heart failure	CC 80		1.28 (1.22, 1.35)
Congenital cardiac/circulatory defects	CC 87-88		0.68 (0.58, 0.79)
Specified heart arrhythmias	CC 92		1.39 (1.33, 1.46)
Precerebral arterial occlusion and transient cerebral ischemia	CC 97		0.87 (0.82, 0.92)
Cerebral atherosclerosis and aneurysm	CC 98		0.84 (0.79, 0.90)
Metastatic cancer and acute leukemia and other major cancers	CC 7-8		2.91 (2.66, 3.18)
Protein-calorie malnutrition	CC 21		1.75 (1.63, 1.88)
Other significant endocrine and metabolic disorders	CC 22-24		0.69 (0.65, 0.74)
Disorders of the vertebrae and spinal discs	CC 39		0.90 (0.86, 0.95)
Other gastrointestinal disorders	CC 36		0.85 (0.81, 0.90)
Other musculoskeletal and connective tissue disorders	CC 43		0.89 (0.85, 0.93)

Iron deficiency and other/unspecified anemia and blood disease	CC 47		1.24 (1.19, 1.30)
Dementia or other specified brain disorders	CC 49-50		1.39 (1.33, 1.45)
Pneumonia	CC 111-113		1.36 (1.29, 1.43)
Decubitus ulcer of skin	CC 148		1.38(1.24, 1.53)

*** Model included both the linear and quadratic versions in the model if the quadratic version of the clinical variable reached the 90% selected rate.**

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

Variation in prevalence of the factor across measured entities

The prevalence of SES factors and African-American patients in the Stroke cohort varies across hospitals. The median percentage of dual eligible patients is 9.5% (interquartile range [IQR] 4.1%-15.9%). The median percentage of African-American patients is 4.2% (IQR 0.0%-14.3%). The median percentage of patients with an AHRQ SES index score equal to or below 47.4 is 20.0% (IQR 3.6%-45.0%).

Empirical association with the outcome (univariate)

The patient-level observed stroke mortality rate is somewhat higher for dual eligible patients, 15.5%, compared with 14.1% for all other patients. The mortality rate for African-American patients was lower at 10.0% compared with 14.8% for patients of all other races. The mortality rate for patients with an AHRQ SES index score equal to or below 47.4 was slightly lower at 14.0% compared with 14.4% for patients with an AHRQ SES index score above 47.4.

Incremental effect of SES variables and race in a multivariable model

We then examined the strength and significance of the SES variables and race in the context of a multivariable model. When we included these variables in a multivariate model that included all of the claims-based clinical variables, only African American race was significantly associated with a lower risk of mortality, with an odds ratio of 0.62. Neither dual eligible or low AHRQ SES index were significant in the multivariable model. In all cases the c-statistics for the stroke patient-level multivariate models with the SDS variables in the models were essentially unchanged from those without (model with original variables: 0.8176; model with dual eligible variable: 0.8176; model with race variable: 0.8184; model with AHRQ SES index variable: 0.8176).

To further understand the relative importance of these risk factors in the measure we compared hospital performance with and without the addition of each SDS variable. We found that the addition of any of these variables into the model had little to no effect on hospital performance. The mean absolute change in hospitals' RSMRs when adding a dual eligibility indicator was 0.00006% with a correlation coefficient between RSMRs for each hospital with and without dual eligibility added of 0.9999. The mean absolute change in hospitals' RSMRs when adding a race indicator was -0.00064% with a correlation coefficient between RSMRs for each hospital with and without race added of 0.9906. The mean absolute change in hospitals' RSMRs when adding a low SES AHRQ indicator was 0.00009% with a correlation coefficient between RSMRs for each hospital with and without low SES added of 0.9999.

Overall we find that the SES variables that could be feasibly incorporated into this model do not have a significant relationship with the outcome in multivariable modeling. For race the relationship with mortality was in the opposite direction than what has been the expressed concern of stakeholders interested in adding such adjustment to the models. We also find that the impact of any of these indicators is very small to negligible on model performance and hospital profiling. Given the controversial nature of incorporating such variables into a risk-model we do not support doing so in a case that is unlikely to affect hospital profiling.

Given these findings and complex pathways that could explain any relationship between SES or race and mortality, which do not all support risk-adjustment, we did not incorporate SES variables and race into the measure.

Future reevaluation efforts will explore the relationship between SES or race and stroke once ICD-10 data are available.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Approach to assessing model performance

During measure development, we computed three summary statistics for assessing model performance (Harrell and Shih, 2001) for the development and validation cohort:

Discrimination Statistics:

(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome.

(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk patients from low-risk patients. Therefore, we would hope to see a wide range between the lowest decile and highest decile)

(3) R-squared indicates how well data fit a statistical model, or the percent of variance explained by the model.

Calibration Statistics:

(4) Over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients)

We tested the performance of the model developed in a randomly selected 50% sample of the hospitalizations for ischemic stroke in **Dataset 1** (development dataset; July 2011-June 2014) by comparing results with those from the validation sample (the remaining 50% of the dataset).

References:

Harrell FE, Shih YCT. Using full probability models to compute probabilities of actual interest to decision makers. *Int. J. Technol. Assess. Health Care* 17 (2001), pp. 17–26.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

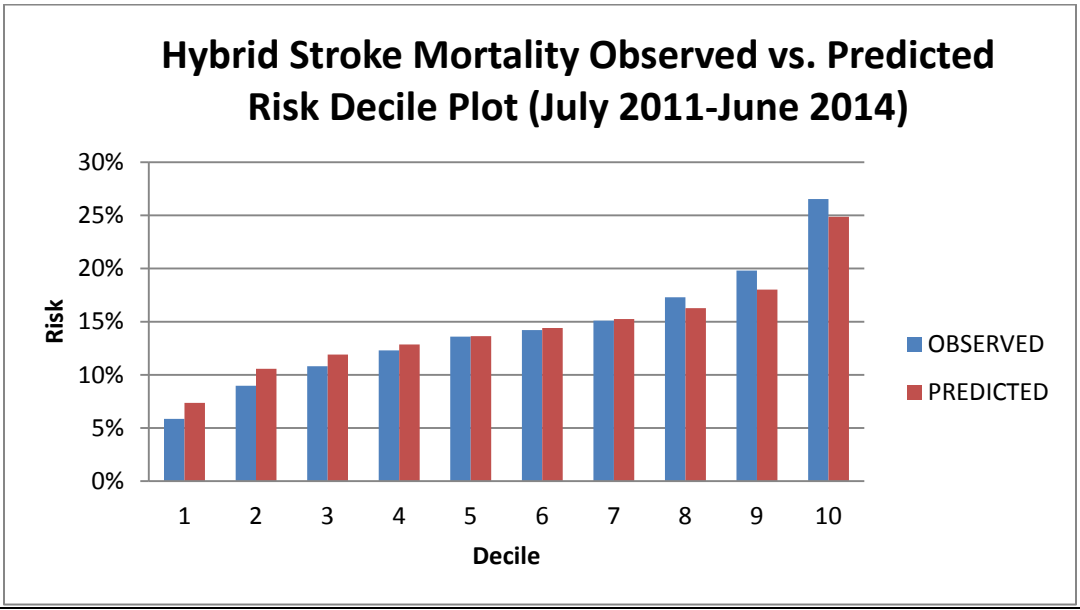
For the development cohort (**Dataset 1**), the results are summarized below:
• 1st half of randomly split sample (development sample): C-statistic = 0.8176; Predictive ability (lowest decile %, highest decile %) = (1.31, 51.05); Adjusted R-squared = 0.2794
• 2nd half of randomly split sample (validation sample): C-statistic = 0.8206; Predictive ability (lowest decile %, highest decile %) = (1.05, 51.93); Adjusted R-squared = 0.2851

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

For the development cohort (**Dataset 1**), the results are summarized below:
- 1st half of randomly split sample (development sample) Calibration: (0.0000, 1.0000)
- 2nd half of randomly split sample (validation sample) Calibration: (0.0000, 1.0000)

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

The risk decile plot is a graphical depiction of the observed mortality in the deciles of the predicted mortality to measure predictive ability. Below, we present the risk decile plot showing the distributions for the development dataset (**Dataset 1**). The plot for the validation dataset was similar.



2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

Discrimination Statistics
The C-statistics of 0.8176 in the development sample and 0.8206 in the validation sample indicate good model discrimination (**Dataset 1**). The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics
Over-fitting (Calibration γ0, γ1)

If the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.

Risk Decile Plots
Higher deciles of the predicted outcomes are associated with higher observed outcomes, which shows a good calibration of the model. This plot indicates excellent discrimination of the model and good predictive ability.

Overall Interpretation
Interpreted together, our diagnostic results demonstrate the hybrid risk-adjustment model adequately controls for differences in patient characteristics (case mix).

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

In addition to the hybrid stroke mortality risk model, we developed a risk model using clinical data (from registry for development but intended to be implemented with electronic clinical data) and no claims data for risk adjustment. Developing a model using this alternate approach to risk adjustment allowed us to examine the performance of a model that could be eventually calculated using EHR data only and would not require claims data.
To develop this model, we used the same merged Medicare claims and registry dataset (**Dataset 1**) as was used to develop the hybrid stroke mortality measure. The same cohort definition that was used for the hybrid stroke mortality measure was also used for the alternative clinical-only model including the ICD-9 condition codes for ischemic stroke, and the inclusion and exclusion criteria. In addition, the same statistical modeling and variable selection approach was used for calculating measure results, except we did not consider any claims-based variables as candidate risk adjustment variables.
To select candidate variables for the model that uses electronic clinical data only for risk adjustment (clinical-only risk model), we began with the same list of variables found in the GWTG-Stroke Registry deemed feasibly extractable from the EHR. The final risk-adjustment variables were selected by a team of clinicians and analysts, primarily based on their clinical relevance, but with knowledge of their strength of association with the mortality outcome. The final clinical-only risk-adjusted model included 9 discrete variables (The quadratic terms of 5 of the linear electronic clinical variables were retained in the model).

Final Model Variables: Model with Registry Data Only for Risk Adjustment (variables meeting criteria in field 2b4.3)

Variable Description	07/2011 – 06/2014 Coefficients (SE)	07/2011 – 06/2014 OR (95% CI)
Age (continuous, per 5 units)	0.0601 (0.0018)	1.37 (1.35, 1.39)
Male	0.2122 (0.0305)	1.21 (1.15, 1.27)
NIH Stroke Scale score (continuous, per 5 units)	0.5092 (0.0075)	1.60 (1.58, 1.62)
Blood glucose/10 (mg/dL): Linear	0.0494 (0.0065)	--
Blood glucose/10 (mg/dL): Square	-0.0008 (0.0001)	--
INR	0.1867 (0.0300)	1.19 (1.13, 1.25)
Heart rate: Linear	-0.0018 (0.0054)	--

Heart rate: Square	0.0001 (0.0000)	--
Weight: Linear	-0.0205 (0.0036)	--
Weight: Square	0.0001 (0.0000)	--
SBP: Linear	-0.0231 (0.0048)	--
SBP: Square	0.0001 (0.0000)	--
DBP: Linear	-0.0216 (0.0039)	--
DBP: Square	0.0001 (0.0000)	--

For model variables that are included in both linear and square forms, odds ratios and 95% CIs are not displayed in the table. The odds ratio for a one unit increase in x is exp(β1+ β2(2x+1)), where β1 is the estimate of the linear term, and β2 is the estimate of the square term. The formula still contains x, so it is not a constant across x, but a function of x.

Approach to assessing model performance

During measure development, we used the same split sample approach using the development and validation samples from Dataset 1. We computed the same three summary statistics for assessing model performance as were used in the development of the hybrid stroke mortality measure (Harrell and Shih, 2001) for the development and validation cohort:

Discrimination Statistics:

(1) Area under the receiver operating characteristic (ROC) curve (the c-statistic) is the probability that predicting the outcome is better than chance, which is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome.

(2) Predictive ability (discrimination in predictive ability measures the ability to distinguish high-risk patients from low-risk patients. Therefore, we would hope to see a wide range between the lowest decile and highest decile)

(3) R-squared indicates how well data fit a statistical model, or the percent of variance explained by the model.

Calibration Statistics:

(4) Over-fitting indices (over-fitting refers to the phenomenon in which a model accurately describes the relationship between predictive variables and outcome in the development dataset but fails to provide valid predictions in new patients)

We tested the performance of the model developed in a randomly selected 50% sample of the hospitalizations for ischemic stroke in **Dataset 1** (development dataset; July 2011-June 2014) by comparing results with those from the validation sample (the remaining 50% of the dataset).

References:

Harrell FE, Shih YCT. Using full probability models to compute probabilities of actual interest to decision makers. *Int. J. Technol. Assess. Health Care* 17 (2001), pp. 17–26.

Results:

Clinical-Only Risk Model (registry clinical data only)

For the development cohort (**Dataset 1**), the results are summarized below:

• 1st half of randomly split sample (development sample): C-statistic = 0.7939; Predictive ability (lowest decile %, highest decile %) = (2.15, 49.45); Adjusted R-squared = 0.2442

• 2nd half of randomly split sample (validation sample): C-statistic = 0.7987; Predictive ability (lowest decile %, highest decile %) = (1.89, 50.93); Adjusted R-squared = 0.2523

For comparison of model with and without inclusion of SES factors, see above section.

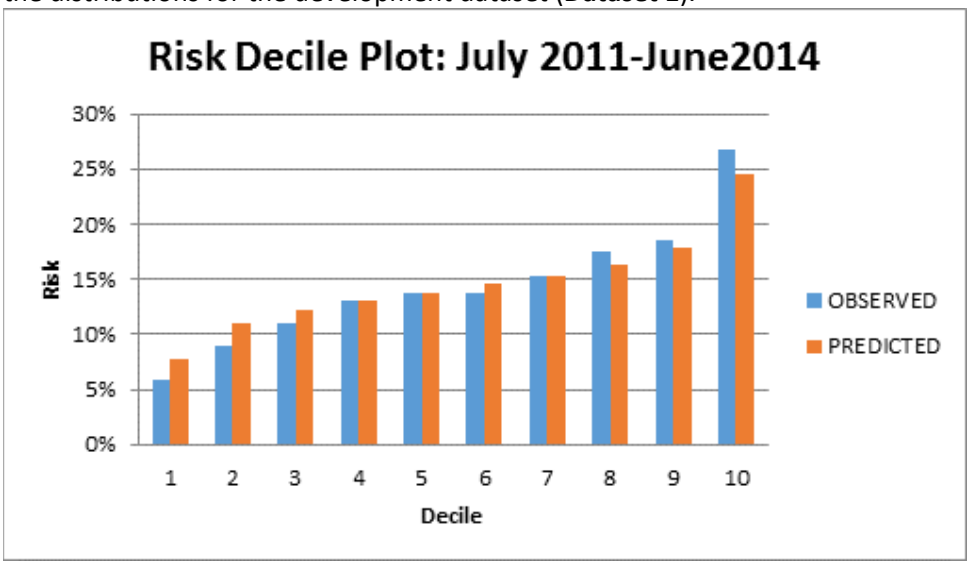
Clinical-Only Risk Model (registry clinical data only)

For the development cohort (**Dataset 1**), the results are summarized below:

- 1st half of randomly split sample (development sample) Calibration: (0.0000, 1.0000)
- 2nd half of randomly split sample (validation sample) Calibration: (0.0000, 1.0000)

Clinical-Only Risk Model (electronic clinical data only)

The clinical-only risk model risk decile plot is a graphical depiction of the observed mortality in the deciles of the predicted mortality to measure predictive ability. Below, we present the clinical-only risk model risk decile plot showing the distributions for the development dataset (**Dataset 1**).



The plot for the clinical-only risk model validation dataset was similar.

Interpretation of Results

Discrimination Statistics

The C-statistics of 0.7939 in the development sample and 0.7987 in the validation sample indicate good model discrimination. The model indicated a wide range between the lowest decile and highest decile, indicating the ability to distinguish high-risk subjects from low-risk subjects.

Calibration Statistics

Over-fitting (Calibration γ0, γ1)

If the γ0 in the validation samples are substantially far from zero and the γ1 is substantially far from one, there is potential evidence of over-fitting. The calibration value of close to 0 at one end and close to 1 to the other end indicates good calibration of the model.

Risk Decile Plots

Higher deciles of the predicted outcomes are associated with higher observed outcomes, which show a good calibration of the model. This plot indicates excellent discrimination of the model and good predictive ability.

Overall Interpretation

Interpreted together, our diagnostic results demonstrate the electronic clinical-only risk-adjustment model adequately controls for differences in patient characteristics (case mix).

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

The method for discriminating hospital performance has not been determined. For public reporting of measures of hospital outcomes developed with similar methodology, CMS characterizes the uncertainty associated with the RSMR by estimating the 95% interval estimate. This is similar to a 95% confidence interval but is calculated differently. If the RSMR's interval estimate does not include the national observed mortality rate (is lower or higher than the rate), then CMS is confident that the hospital's RSMR is different from the national rate, and describes the hospital on the Hospital Compare website as "better than the U.S. national rate" or "worse than the U.S. national rate." If the interval includes the national rate, then CMS describes the hospital's RSMR as "no different than the U.S. national rate" or "the difference is uncertain." CMS does not classify performance for hospitals that have fewer than 25 cases in the three-year period.

However, the measure is not currently publicly reported, and decisions about the approach to discriminating hospital performance have not been made.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Analyses of Medicare FFS data show substantial variation in RSMRs among hospitals. Using data from the development sample (**Dataset 1**, July 2011-June 2014), the median hospital RSMR for the hybrid model was 14.50% with a range of 10.67% to 19.15%. The interquartile range was 13.62% - 15.07%.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

Despite recent decreases in mortality rates nationally, stroke is the fifth most common cause of death in the United States, affecting approximately 795,000 people annually, and has a 30-day mortality rate that varies by age from 9% in patients 65 to 74 years of age, 13.1% in those 74 to 84 years of age, and 23% in those ≥85 years of age [Go et al., 2014; Kochanek et al., 2014; Casper et al., 2008].

The variation in RSMRs suggests that there are differences in the quality of care received across hospitals for stroke that support measurement to reduce this variation.

References:

Casper ML, Nwaise IA, Croft JB, Nilasena DS. Atlas of Stroke Hospitalizations Among Medicare Beneficiaries. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention; 2008.

Go AS, Mozaffarian D, Roger VL, et al. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. *Circulation.* Jan 21 2014;129(3):e28-e292.

Kochanek KD MS, Xu JQ, Arias E. Mortality in the United States, 2013. NCHS data brief, no 178. 2014.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note*: This criterion is directed to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***If comparability is not demonstrated, the different specifications should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to demonstrate comparability of performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Note that in section 2b.4.11 we describe a second approach to risk adjustment for this measure that uses only clinical data elements from patient EHRs and we provide results of the analysis of the performance of that alternate risk model. We have also submitted, as a separate measure, a claims-only stroke mortality measure for endorsement. We compare the performance of all three risk models, the hybrid model, the EHR-only alternate model, and the claims-only model, in 2b2.3 and 2b2.4.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Results provided in 2b2.3

2b6.3. What is your interpretation of the results in terms of demonstrating comparability of performance measure scores for the same entities across the different data sources/specifications? (i*.e., what do the results mean and what are the norms for the test conducted*)

Interpretation provided in 2b2.4

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Since the Medicare claims data that we receive does not contain missing data, we only conducted testing to identify the extent and distribution of missing clinical data from the GWTG-Stroke Registry. We performed this testing to demonstrate that results are not biased due to systematic missing data. Then, to deal with the missing data, we used multiple imputation to predict and impute the missing variables based on other non-missing patient variables.

There is no current national database that includes NIH Stroke Scale score data for stroke patients admitted to all non-federal acute care hospitals. Therefore, Implementation of this measure depends on hospitals including NIH stroke severity scores for all patients admitted with ischemic stroke, in the claims they submit to Medicare using the ICD-10 codes that will be available beginning in October 2016. Once these ICD-10 codes become available, it will be important to assess rates of missing data. Collection of the NIH Stroke Scale is now Class I recommended in the AHA/ASA guidelines for care of patients admitted with acute ischemic stroke. Given the importance of this measure in national reporting programs and the importance of stroke severity to this measure's risk model, CMS expects high rates of NIH Stroke Scale reporting and will determine the plan for handling missing data during measure implementation.

There is also no current mechanism or mandate for hospitals to report the CCDE or the critical data elements from the CCDE used in the hybrid stroke mortality measure. Therefore, implementation of this measure will depend on implementation of a national reporting mechanism for the critical EHR data elements.

GWTG-Stroke Registry Data Element: NIH Stroke Scale Score

For measure development we assessed the frequency of missing NIH Stroke Scale scores from the GWTG-Stroke Registry during the development of the measure. The registry data on stroke severity was used as a surrogate for data that will eventually come either from the claims, once the ICD-10 codes for stroke severity are available and consistently used by hospitals, or from the EHR if the data element can be shown to be feasible for extraction and reporting. Among the measure cohort, NIH Stroke Scale scores were available in 82.39% of patients with an admission for ischemic stroke from July 1, 2013 and June 30, 2014. This proportion increased from 70.78% of admissions occurring between July 1, 2011 and June 30, 2012. Using the registry stroke severity data, we used multiple imputation to produce estimates for our testing and analyses. The multiple imputation technique used to impute missing values was a multi-logit regression model. Five copies of imputation datasets were produced for the analyses. The results based on these data were aggregated according to the standard statistical methods for the presentation of the results and for the measure score calculation.

In multiple imputation, missing variable values are predicted using other related patient variables available. The predicted values are substituted for the missing values, which results in a full data set without any missing variables (the imputed data set). By repeating this process multiple times, we get multiple imputed data sets. We then conduct analyses on and obtain results for each imputed data set. The results based on multiple data sets are combined to produce the overall final results. The multiple imputation represents a random sample of the missing values according to the association of the non-missing values of all the variables considered, and the resulting inferences of multiple imputation are statistically valid, which reflect the uncertainty due to missing values (He et al., 2014).

To determine if patterns of missing stroke severity data were systematic among hospitals participating in the GWTG-Stroke Registry in a way that might bias measure results, we examined the correlation between the rate of missing NIH Stroke Scale among patients and hospital-level 30-day mortality rate for all admissions, those with NIH Stroke Scale scores and those missing NIH Stroke Scale scores. We also examined the correlation among only those admissions with missing NIH Stroke Scale scores. While this statistical measure was developed using multiple imputation to account for missing NIH Stroke Scale data, approaches to handling missing NIH Stroke Scale data in measure calculation will be reassessed during implementation.

References:

He R, Belin T. Multiple imputation for high-dimensional mixed incomplete continuous and binary data. Stat. Med. 2014;33:2251–2262.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

In the development dataset which used registry data as a surrogate for future claims data on stroke severity, NIH Stroke Scale scores were available in 82.39% of patients with an admission for ischemic stroke from July 1, 2013 and June 30, 2014. This proportion increased from 70.78% of admissions occurring between July 1, 2011 and June 30, 2012. For all 3 years of data combined, the NIH Stroke Scale score was missing in 23.51% of all patients (N=188,975). At the hospital level, the median percentage of missing NIH Stroke Scale scores was 10.15% (interquartile range 3.89%-27.27%), with a

range of 0% to 100%, between July 1, 2013 and June 30, 2014. This percentage improved from July 1, 2011 to June 30, 2012, when the median percentage of missing NIH Stroke Scale scores was 19.23% (interquartile range 7.15%-50.00%), with a range of 0% to 100%. For all 3 years of data combined, the median percentage of missing NIH Stroke Scale scores at the hospital level was 15.89% with a range of 0% to 100%. The interquartile range was 6.74% to 39.27%.

The correlation between the rate of missing NIH Stroke Scale scores in the registry dataset and the hospital-level 30-day mortality rate was not significant overall. In the subset of patients who had missing NIH Stroke Scale scores, the correlation between the rate of missing NIH Stroke Scale scores and the hospital-level 30-day mortality rate was -0.2568 ($p<0.0001$).

The percent of final risk-adjustment variables missing in the registry data included in the development dataset (N=94,466) are as follows:

Variable	**Percent missing**
NIH Stroke Scale	23.49
Blood glucose (mg/dL)	20.96
Vital Signs - Heart Rate	19.96
Vital Signs - Blood Pressure Systolic	19.06
Vital Signs - Blood Pressure Diastolic	19.11

We did not include any variable as a candidate variable that had more than 30% of its values missing from **Dataset 1**. For example, we anticipated that initial platelet count at hospital arrival would be an important predictor of mortality to include in the hybrid model; however, it was missing for 99.97% of the patients in the development dataset, and was therefore not included as a candidate variable.

Variables above were only missing from **Dataset 1**, which was GWTG-Stroke Registry data. Hospitals across the United States voluntarily participate in the GWTG-Stroke Registry and are not required to report all variables, based on our testing of the CCDE we expect a lower rate of missing among hospitals when the clinical data elements are extracted from the EHRs for implementation.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

The overall rate of missing NIH Stroke Scale was relatively low. The results are not biased due to systematic missing data (or differences between patients with reported versus unreported data), as we used multiple imputation to substitute missing values with predicted values. The proportion of patients missing NIH Stroke Scale scores had little impact on hospital 30-day mortality rate. However, the stroke severity score and other clinical data from the GWTG-Stroke Registry are only used as a surrogate for data that will eventually come from Medicare claims and hospital EHRs. These analyses will need to be repeated in the future using a claims and EHR merged dataset.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields?** *(i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields)*
ALL data elements are in defined fields in a combination of electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.**

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL.**
Attachment **Attachment:** NQF_2877_Hybrid_Stroke_Mortality_Feasibility_Scorecard_v1.1.docx

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PROM data (patients, service recipients, respondents) and those whose performance is being measured.
Administrative data are routinely collected as part of the billing process. New ICD-10 codes for NIH Stroke Scale scores will be available to hospitals to include in Medicare claims in October 2016.

Electronic clinical data will be collected from hospitals using MAT output and value sets to inform data queries and electronic reporting requirements.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
There are no fees associated with the use of this measure.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Planned	Current Use (for current use provide URL)
Public Reporting Not in use	

4a.1. For each CURRENT use, checked above, provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included

N/A

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

This is a new measure

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

CMS intends to implement this measure in the Hospital Inpatient Quality Reporting (HIQR) Program once the new NIH Stroke Scale ICD-10 codes have been in use and the core clinical data elements (CCDE) have been reported by hospitals for 3 years. This measure requires 3 years of data for calculation. This timeline depends on the consistent capture of NIH Stroke Scale and use of the new ICD-10 codes by hospitals for all patients admitted with acute ischemic stroke as is recommended in the current AHS/ASA clinical guidelines. In addition, it requires implementation of a reporting mechanism for the CCDE. Once this new measure is implemented it will replace other Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure.

4b. Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b.1. Progress on Improvement. (Not required for initial endorsement unless available.)

Performance results on this measure (current and over time) should be provided in 1b.2 and 1b.4. Discuss:

- **Progress (trends in performance results, number and percentage of people receiving high-quality healthcare)**
- **Geographic area and number and percentage of accountable entities and patients included**

This is a new measure and there is no information available on performance improvement. However, there has been significant progress in 30-day RSMR for stroke when looking at the current publicly reported measure. The median 30-day RSMR decreased by 1.1 absolute percentage points from 2011-2012 (median RSMR: 15.3%) to 2013-2014 (median RSMR: 14.2%). The median hospital RSMR from 2011-2014 was 14.9% (IQR 14.2% - 15.6%).

4b.2. If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of

high-quality, efficient healthcare for individuals or populations.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Were any unintended negative consequences to individuals or populations identified during testing; OR has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them.**

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

0467 : Acute Stroke Mortality Rate (IQI 17)

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

Title: Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure

Steward: Centers for Medicare & Medicaid Services (CMS)

5a. Harmonization

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**

Are the measure specifications completely harmonized?

No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

We did not include in our list of related measures any non-outcome (such as process) measures with the same target population as our measure. Because this is an outcome measure, clinical coherence of the cohort takes precedence over alignment with related non-outcome measures. Furthermore, non-outcome measures are limited due to broader patient exclusions. This is because they typically only include a specific subset of patients who are eligible for that measure (for example, patients who receive a specific medication or undergo a specific procedure). Additionally, this measure and the NQF Acute Stroke Mortality Rate (IQI 17) (AHRQ) Measure #0467 are complementary and related rather than competing measures. Although they both assess mortality for patients admitted to acute care hospitals with a principal discharge diagnosis of acute ischemic stroke, the specified outcomes are different. Our measure assesses 30-day mortality while #0467 assesses inpatient mortality. The 30-day mortality and inpatient mortality outcomes each have distinct advantages and uses, which make them complementary (and related) as opposed to competing. For example the 30-day period provides a broader perspective on hospital care and utilizes a standard time period to examine hospital performance to avoid bias by differences in length of stay among hospitals. However, in some settings it may not be feasible to capture post-discharge mortality, making the inpatient measure more useable. We have previously consulted with AHRQ to examine

harmonization of the measures' cohort. As a result of that collaboration, we have found that the measures' cohorts are harmonized to the extent possible and that the small differences in cohort inclusion and exclusion criteria are appropriate because the measures assess different outcomes. The NQF Acute Stroke Mortality Rate (IQI 17) (AHRQ) Measure #0467 is also intended for patients 18 years of age and older, which represents a different cohort than the 65 and older Medicare population for this new hybrid measure.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
This measure looks at a longer outcome time frame (30-days versus in-hospital) and incorporates stroke severity into the risk-model.

The current publicly reported measure, Hospital 30-Day Mortality Following Acute Ischemic Stroke Hospitalization Measure, is a potentially competing measure. It is CMS intent to replace the current measure in any given program with this newly developed measure, which includes stroke severity in the risk model.

The Claims-based 30-Day, All-Cause, Risk-Standardized Mortality Rate (RSMR) Following Acute Ischemic Stroke with Risk Adjustment for Stroke Severity measure is also being submitted to NQF for endorsement. This measure uses only claims but is otherwise harmonized with this new hybrid measure. It is CMS intent to implement only one of the new stroke mortality measures (this hybrid measure or the claims-only measure) in any given program.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Claims-Only_-_Hybrid_Stroke_Mortality_Measure_Tech_Report_1-15-16.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Centers for Medicare & Medicaid Services (CMS)
Co.2 **Point of Contact:** Lein, Han, Lein.han@cms.hhs.gov, 410-786-0205-
Co.3 **Measure Developer if different from Measure Steward:** Yale New Haven Health Services Corporation – Center for Outcomes Research and Evaluation (CORE)
Co.4 **Point of Contact:** Karen, Dorsey, Karen.dorsey@yale.edu, 203-764-5700-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Our working group consisted of the following members:
- Lee Schwamm, MD: Vice Chairman, Department of Neurology, Massachusetts General Hospital
- Gregg Fonarow, MD: Professor of Medicine, University of California, Los Angeles
- Jason Sico, MD, FACP: Director, Stroke Care VA Connecticut Healthcare System
- Kevin Sheth, MD: Associate Professor of Neurology and Neurosurgery, Yale University.

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.2 **Year the measure was first released:**
Ad.3 **Month and Year of most recent revision:**
Ad.4 **What is your frequency for review/update of this measure?** N/A

Ad.5 **When is the next scheduled review/update for this measure?**
Ad.6 **Copyright statement:** N/A Ad.7 **Disclaimers:** N/A
Ad.8 **Additional Information/Comments:**