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 #: 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 and 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 (OQR) 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).

REFERENCES:

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) 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.

Numerator Statement: Emergency department (ED) acute ischemic stroke or hemorrhagic stroke patients arriving at the ED within 2 hours of the time last known well, with an order for a head CT or MRI scan whose time from ED arrival to interpretation of the Head CT scan is within 45 minutes of arrival.

Denominator Statement: Emergency department acute ischemic stroke or hemorrhagic stroke patients arriving at the ED within two hours of the time last known well with an order for a head CT or MRI scan.

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.

Measure Type: Process

Data Source: Claims, Electronic Health Records, Paper Medical Records

Level of Analysis: Facility, Other

IF Endorsement Maintenance – Original Endorsement Date: Jan 17, 2011 **Most Recent Endorsement Date:** Sep 23, 2016

Preliminary Analysis: Maintenance of Endorsement

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 *structure, 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. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

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 2016**

- 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.
- Guideline #2 provides a focused update of the current recommendations for the endovascular treatment of acute ischemic stroke.
- The developer stated 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

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 three guidelines (2 from previous submission and 1 new) from the American Heart Association/American Stroke Association for the early management of patients with acute ischemic stroke including endovascular treatment
- Guideline 1 Recommendations: American Heart Association/ American Stroke Association
    - 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)**.
    - 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).**
    - 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)**.
- Guideline #2 Recommendation: American Heart Association/ American Stroke Association
    - 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)**.
- Guideline #3 Recommendation: American Heart Association/ American Stroke Association

- All patients admitted to hospital with suspected acute stroke should receive brain imaging evaluation on arrival to hospital. In most cases, noncontrast CT (NCCT) will provide the necessary information to make decisions about acute management. **(Recommendation revised from 2013 guideline). (Class I, Level of Evidence B Nonrandomized**).
- Systems should be established so that brain imaging studies can be performed within 20 minutes of arrival in the ED in at least 50% of patients who may be candidates for IV alteplase and/or mechanical thrombectomy. **(New recommendation). (Class I, Level of Evidence B Nonrandomized)**.

- The developer also summarized eight articles, which they state support the measure's intent; however, the articles largely 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
Not applicable

Questions for the Committee:

- *Does the Committee believe that the evidence submitted for the 45-minute window is adequate?*
- *The evidence provided by the developer is updated, directionally the same, and stronger compared to that for the previous NQF review. Does the Committee agree there is no need for repeat discussion and vote on Evidence?*
- *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?*
- *If derived from patient report, does the target population value the measured process or structure and find it meaningful?*

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

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.

- Performance gap has narrowed, but a gap still exists from previous endorsement.

- The developer provided the following data for current performance:

	2012-2013	2017-2018
# facilities	918	1,550
# CT/MRIs performed (den)	16,817	3,1939
Weighted mean performance	59.6%	75.0%
Standard deviation	99.9	82.2
Facility median	62.0%	79.0%
Range	0 – 100%	0 -100%
10th percentile	24.0%	46.0%
25th percentile	42.0%	64.0%
90th percentile	88.0%	94.0%

Disparities

- The developer indicated that 2014 data showed 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.
- However, more recent data (2016 – 2018) showed no race disparities for head CT or MRI scan interpreted within 45 minutes of ED arrival; however, Hispanic patients remained less likely to be included in the measure's numerator compared to non-Hispanic patients. Females also remained less likely than males 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:

- *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)

Evidence

- **There is good evidence relating percent of carotid stenosis to appropriate intervention decision. The numerator calls for a measurement of carotid diameter; however, the rationale discusses that it is inconsistency in the calculation of stenosis that is problematic. It appears that NASCET is preferable to ECST for this calculation. It is not clear whether the mention of a number in the numerator will standardize that unless the report mentions the method of calculation.**
- **The evidence presented supports an association between onset of symptoms and effectiveness of intervention. The process being examined is directly related to desired outcomes, as there is a clear and well-supported association between early intervention (which is contingent on early scan and interpretation) and more favorable clinical outcomes.**
- **Applies directly.**
- **The developer has now provided three guidelines from the American Heart Association/American Stroke Association for the early management of patients with acute ischemic stroke including endovascular treatment. The developer stated 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. However, I would like to see data on the relationship of this measure to patient outcomes. I am not aware of any new studies/information that changes the evidence base for this measure.**
- **This measure is an old measure and I am aware of no new data.**
- **There is very strong evidence that earlier treatment with tPA is valuable. Some sort of CT or MRI interpretation has to happen prior to that time. So, there is a very strong argument for the principle of this measure. The 45 minute time window is arbitrary and there is little evidence supporting that specific threshold. However, given that a gap persists even at that threshold, there is a reasonable argument that the measure is important and worth reporting.**

Performance Gap

- **Seems that the measure is improving performance.**
- **More recent data indicates that gaps in care exist regarding this process measure, particularly regarding Hispanic patients, female patients, and patients in smaller facilities (<50 beds) or major teaching facilities.**
- **Unclear i.e. level of disparity.**
- **There still remains a performance gap with Hispanics and females, therefore it warrants a national performance measure. An additional gap in care I would like to consider is with patients under 18. Was there a reason this population was not included? The 2019 "Management of Stroke in Neonates and Children: A Scientific Statement From the American Heart Association/American Stroke Association" states, "Median time to radiological confirmation of diagnosis is 15 to 24 hours. Children with onset of stroke during admission for other illnesses experience similar delays in radiological confirmation of ischemic stroke. The major causes of delays include delayed consideration of stroke among frontline providers and delays in accessing MRI, often related to the need for sedation or anesthesia. Delays are greater in evenings and weekends." https://www.ahajournals.org/doi/10.1161/STR.0000000000000183**

- **Yes, the performance gap exists but is moving, disparities are related to rural location and teaching institutions which may be markers for SES.**
- **Yes. A meaningful gap persists.**

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

- **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/facility level of analysis

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 – less emphasis if no new testing data provided.

For maintenance measures, summarize the reliability testing from the prior review:

- 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.
- Data element validity testing was performed and counted for data element reliability as well – see validity testing section.

Updates to testing:

- The dataset used for testing included cases submitted from 1,550 facilities to Hospital Compare from July 1, 2017 – June 30, 2018. The sample included 31,939 denominator cases (initial population) and 23,953 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.
- Reliability scores ranged from 0.52 to 1.00. The median reliability score was 0.76. A value of 0.7 or higher is often regarded as acceptable reliability

Validity

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 – less emphasis if no new testing data provided.

Complex measure evaluated by Scientific Methods Panel? ☐ **Yes** ☒ **No**

- Empirical validity of critical data elements was assessed by examining kappa statistics (for categorical variables and the constructed outcomes of the numerator and denominator) and Pearson's correlation coefficient (for non-categorical variables) between facility abstraction and auditor (CDAC) abstraction for each of the data elements used to calculate the measure.
- The analysis used data elements for 2,622 cases abstracted by CDAC, which were previously abstracted by facilities; these data were collected from July 1, 2016 to December 31, 2018.
- The developer provides validity testing for the eight data elements. The agreement between facility and CDAC-abstracted data elements ranged from moderate to strong across the data elements. Kappa values ranged from 0.77–0.93 for categorical data elements; Pearson's correlation coefficients for non-categorical variables ranged from 0.51-0.92; and Kappa values for the constructed variables of the numerator and denominator were each 0.85. These results support the validity of the measure and its calculation.

2b2-2b6. Potential threats to validity should be assessed/addressed.

2b2. 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		10,702	3.94	2.0	4.0	7.0
Head CT Scan or MRI Order	X		7,635	2.92	1.0	2.0	5.0
Last Known Well	X		101,127	37.21	27.0	44.0	69.0
Date Last Known Well		X	0	0.00	0.0	0.0	0.0
Time Last Known Well		X	192	0.00	0.0	0.0	0.0
Arrival Time		X	0	0.0	0.0	0.0	0.0
Last Known Well Minutes	X		77,757	39.38	22.0	34.0	52.0
Head CT Scan or MRI Interpretation Date		X	0	0.00	0.0	0.0	0.0
Head CT Scan or MRI Interpretation Time		X	54	0.00	0.0	0.0	0.0
Total Denominator Exclusions	4 exclusions	-	197,221	72.66	-	-	-
Total Numerator Exceptions	-	5 exceptions	246	0.00	-	-	-

Data Element	Denominator Exclusion or Numerator Exception?		Overall Occurrence		Distribution Across Facilities (%)		
	Denominator Exclusion	Numerator Exception	N	%	25th	50th	75th
Total Removed from the Denominator or Numerator	9 exceptions and exclusions		197,467	72.66	-	-	-

- Excluded a significant ~73% of population

2b3. Risk adjustment:

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

2b4. Meaningful difference (can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified):

- Using data from July 1 2017 – June 30, 2018, the developer tested the statistical significance of the difference between facility performance scores and the mean performance value for 1,550 facilities meeting public-reporting requirements.
- Results of the analysis indicated that the performance of 5.9% of the 1,550 facilities (n=92) was statistically significantly different from the average performance rates.

2b5. Comparability of data sources/methods:

- According to the developer this measure only uses only one set of specifications

2b6. Missing Data

- Missing data are 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.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *The NQF is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?*

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions)?*
- *Are exlcusions clinically relevant?*
- *The NQF staff is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?*

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

- 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 validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

- Potential threats to validity assessed (Box 1) → empirical validity testing conducted (Box 2) → no empirical testing at the measure score level (Box 5) → validity testing conducted with patient-level data elements (Box 9) → percent agreement of the critical data elements assessed (Box 10) → appropriate method used (Box 11) → Moderate certainty or confidence that the data used in the measure are valid (Box 11a) → Moderate
- (Note: Moderate is the highest rating possible with data element validity only)

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

Reliability-Specifications

- **No concerns.**
- **Data elements are clearly defined, and reliability testing is sound. No concerns about the measure being consistently implemented.**
- **It can be adopted in almost all settings that should be seeing such patients.**
- **The data elements seem to be clearly defined. I do however question why CT and MRI are used interchangeably. In pediatric cases, a CT scan can appear "normal" with ischemic stroke, therefore an MRI is preferable.**
- **No evidence to support measure is worse than expected originally.**
- **Unclear on how "interpretation" is defined. It seems that time to scan is used — does that qualify as interpretation?**

Reliability-Testing

- **No concerns.**
- **The reliability of the measure appears sound.**
- **No concerns.**
- **I believe the exclusions as stated are relevant. However, I would like consider the possibility of including under age 18 as stated above in 4.1b.**
- **No concerns.**
- **No concerns.**

Validity-Testing

- **Face validity only, but good agreement.**
- **No concerns about the testing results.**

- **No concerns.**
- **I do not have any concerns with validity/testing results.**
- **No concerns.**
- **The big issue between facilitiy variation on missiness of last normal time. The magntidue of missingness, on the whole, is enormous. While that's not fully unexpected, the bigger concern is the variation between centers which suggests that this element could be use dto game the system and/or that the actual performance of individual centers could be severely obscured. More to the point, part of providing high quality care for stroke is to put in effort to define the last normal time. Simply saying, "we don't know" and stopping isn't quality care.**

Threats to Validity

- **There is probably a meaningful difference. No information for missing data.**
- **Missing data is appropriate excluded from the process measurement. Exclusion criteria appear clinically appropriate.**
- **Missing data would be a concern.**
- **The developer points out that if data are missing, the case will be rejected. There is moderate certainty or confidence that the performance measure scores are reliable. However, 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.**
- **No new threats to validity.**
- **No concerns.**

Other Threats to Validity

- **No exclusions.**
- **Exclusions are consistent with evidence.**
- **Exclusions would need to be examined.**
- **There are no patient groups that are inappropriately excluded. I assume there was a reason for excluding the pediatric population. There is moderate certainty or confidence that the data used in the measure are valid.**
- **Risk adjustment not necessary and I agree with Sponsor's reason for not adjusting this.**
- **No concerns.**

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

- **No issues with feasibility.**
- **Required data elements are routinely generated during care delivery. No concerns about feasibility of data collection.**
- **Should be available in an EHR.**
- **I don't have concerns about the data collection strategy and how it is used. 80% of participants agreed that practical aspects of reporting this chart-abstracted measure do not place undue burden on facilities that collect the data.**
- **Currently used feasibility documented.**
- **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

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. 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.

4a.1. 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.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

Planned 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.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

- The developer reports that, feedback received from stakeholders (via the RightNow Q&A tool) is used to revise the measure specifications. To date, they have received no significant concerns raised by stakeholders about the measure specifications through their feedback collection tool.
- In addition, stakeholders may submit comments on the measure through the Outpatient Prospective Payment System (OPPS) annual rule-making process. No comments were received for this measure during the Calendar Year (CY) 2016 -2019 OPPS rule-making cycles

Additional Feedback:

- The developer reports that to date, they have received no significant feedback about the measure specifications.

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*
- *How has the measure been vetted in real-world settings by those being measured or others?*

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

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability 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.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

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 27% from 62.0% in 2012 to 79.0% in 2018.

4b2. Benefits vs. harms. 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).

Unexpected findings (positive or negative) during implementation

- The developer reports that they 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.

Potential harms

- The developer reports that they did not identify any unintended consequences from implementing this measure.

Additional Feedback:

- The developer states that no evidence of unintended consequences to individuals or populations has been reported by external stakeholders since its implementation.

***Questions for the Committee**:*

- *How can the performance results be used to further the goal of high-quality, efficient healthcare?*

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

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

Use

- **Good uptake, increased after measure introduced.**
- **Measure is publicly reported through CMS HOQR Program and are publicly available, informing participating hospital comparisons. No significant feedback from stakeholders has been received suggesting substantive changes to the measure. The results of this measure can be used by facilities of varying size and staffing profiles to expedite time to appropriate treatment intervention.**
- **Unclear about accountability.**
- **The measure is being publicly reported through the Hospital OQR Program, which is a pay for quality data reporting program implemented by CMS for outpatient hospital services. It is available to the public through the CMS Hospital Compare website. Feedback received from stakeholders (via the RightNow Q&A tool) is used to revise the measure specifications. Following receipt of a suggestion to adjust the specifications, a literature review is performed to determine if the proposed change aligns with the empirical evidence base for the measure; feedback from the expert work group is obtained to evaluate the change to the specifications. To date, they have received no significant concerns raised by stakeholders about the measure specifications through the RightNow Q&A tool.**
- **Public Reporting, HOQR, etc.**
- **No concerns.**

Usability

- **No unintended consequences.**
- **The evidence presented suggests that facilities can use performance on this measure to increase high-quality, efficient health care for the target population. No unintended consequences are expected or noted.**
- **Harms would be low.**
- **The median rate of head CT or MRI scans interpreted within 45 minutes of ED arrival, has increased 27% between 2012 and 2018. Which is substantial, but perhaps with increased efficiency, data reporting and increasing the number of facilities, that number can increase even more. There have not been any unintended consequences reported to individuals or populations by external stakeholders since its implementation.**
- **No identified unintended consequences.**
- **No concerns.**

Criterion 5: Related and Competing Measures

Related or competing measures
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.

Harmonization

- The developer indicated that the specifications are harmonized to the extent possible
- The NQF team does not believe that harmonization is warranted

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

- **No concerns.**
- **There is no need to harmonize with an identified complementary measure (HIQR), which focuses on thrombolytic therapy for inpatient settings.**
- **There is one NQF endorsed measure: 0437 : STK 04: Thrombolytic Therapy. And one non NQF measure: 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). The NQF measure-maintenance teams for both reporting programs meet periodically to resolve any inconsistencies in the interpretation or guidance provided for the shared data elements. The ICSI measure is related to NQF #0661, but it focuses on head CT completion, which is an intermediate step for head CT interpretation (NQF #0661).**
- **Measure harmonizes with other measures.**
- **No concerns.**

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: Month/Day/Year

- **Of the XXX NQF members who have submitted a support/non-support choice:**
  - XX support the measure
  - YY do not support the measure

1. Evidence and Performance Gap – 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 sub criteria 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_-Updated-.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

No

1a. Evidence (subcriterion 1a)

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

- ☐ 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. (A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)
- ☐ 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
  - ☐ Appropriate use measure: Click here to name what is being measured
- ☐ Structure: Click here to name the structure
- ☐ Composite: Click here to name what is being measured

1a.2 **LOGIC MODEL** Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

According to the Centers for Disease Control, stroke is the fifth leading cause of death (Kochanek et al., 2014; Heron, 2018). 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.3**) (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.

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.

Jauch E.C., Saver J.L., Adams H.P. Jr, Bruno A., Connors J.J., Demaerschalk B.M., 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. (2013). 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, 44(3), 870–947.

Heron, Melonie. (2018). Deaths: Leading causes for 2016. *National Vital Statistics Reports*. 67(6). https://www. cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_06.pdf.

1a.3 **Value and Meaningfulness: IF** this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

Not applicable, as this measure is not derived from patient-reported data.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 **FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.**

This measure is not a health outcome/PRO-PM.

1a.3. **SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.**

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☒ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body ofevidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☒ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Three 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. A third guideline was published in 2018 to provide comprehensive recommendations regarding care for patients with acute arterial ischemic stroke. Citations for the three guidelines follow: 1—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. 2—Powers WJ, Derdeyn CP, Biller J, Coffey CS, Hoh BL, Jauch EC, Johnston KC, Johnston SC, Khalessi AA, Kidwell CS, Meschia JF, Ovbiagele B; Yavagal DR; on behalf of the American Heart Association Stroke Council. 2015 AHA/ASA focused update of the 2013 guidelines for the early management of patients with acute ischemic stroke regarding endovascular treatment: A guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke. 2015;46. Guideline available at: http://stroke.ahajournals.org/content/early/2015/06/26/STR.0000000000000074.full.pdf+html. 3—Powers WJ, Rabinstein AA, Ackerson T, Adeoye OM, Bambakidis NC, Becker K, Biller J, Brown M, Damaerschalk BM, Hoh B, Jauch EC, Kidwell CS, Leslie-Mazwi TM, Ovbiagele B, Scott PA, Sheth KN, Southerland AM, Summers DV, Tirschwell DL; on behalf of the American Heart Association Stroke Council. 2018 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. 2018;e46. Guideline available at: https://www.ahajournals.org/doi/pdf/10.1161/STR.0000000000000158.
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being	*Guideline 1* provides recommendations for patients with acute cerebral ischemic symptoms that have not yet resolved. Three recommendations support the measure's clinical intent.

measured. If not a guideline, summarize the conclusions from the SR.	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. 3032). *Guideline 3* provides a comprehensive and updated set of recommendations for patients with acute arterial ischemic stroke. Two of the recommendations support the measure's intent. Guideline #3 *Recommendation:* American Heart Association/ American Stroke Association A. All patients admitted to hospital with suspected acute stroke should receive brain imaging evaluation on arrival to hospital. In most cases, noncontrast CT (NCCT) will provide the necessary information to make decisions about acute management. (Recommendation revised from 2013 guideline). (Class I, Level of Evidence B Nonrandomized; pg.e58). B. Systems should be established so that brain imaging studies can be performed within 20 minutes of arrival in the ED in at least 50% of patients who may be candidates for IV alteplase and/or mechanical thrombectomy. (New recommendation). (Class I, Level of Evidence B Nonrandomized; pg.e58).
Grade assigned to the **evidence** associated with the recommendation with the	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

definition of the grade	(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.
Provide all other grades and definitions from the evidence grading system	**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. **The following grading scale applies to recommendations from *Guideline 3:*** *Recommendation A: Class I, Level B-NR (Nonrandomized):* moderate-quality

	evidence from 1 or more well-designed, well-executed nonrandomized studies, observational studies, or registry studies; meta-analyses of such studies. *Recommendation B: Class I, Level B-NR:* moderate-quality evidence from 1 or more well-designed, well-executed nonrandomized studies, observational studies, or registry studies; meta-analyses of such studies.
Grade assigned to the **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.
Provide all other grades and definitions from the **recommendation** grading system	**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. **The following grading scale applies to recommendations from *Guideline 3:*** *Recommendation A: Class I, Level B-NR:* moderate-quality evidence from 1 or more well-designed, well-executed nonrandomized studies, observational studies, or registry studies; meta-analyses of such studies. *Recommendation B: Class I, Level B-NR:* moderate-quality evidence from 1 or more well-designed, well-executed nonrandomized studies, observational studies, or registry studies; meta-analyses of such studies.
Body of evidence: • **Quantity—how many studies?** • **Quality—what type of studies?**	The three guidelines are evidenced based; details are provided below. 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. 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. Guideline 2 does not indicate the specific number or type of study designs included in the body of evidence; however, the Class I recommendation is Level A evidence and unchanged from the 2013 guideline. Level A evidence is defined as high-quality evidence from more than 1 RCT, meta-analyses of high-quality RCTs, or one or more RCTs corroborated by high-quality registry studies.

	Guideline #3 does not indicate the specific number or type of study designs included in the body of evidence; however, the two Class I recommendations are Level B-NR. Level B-NR evidence is moderate-quality evidence from 1 or more well-designed, well-executed nonrandomized studies, observational studies, or registry studies, or meta-analyses of such studies. The two recommendations referenced 12 unique citations with evidence from systematic reviews, nonrandomized studies, and observational studies.
Estimates of benefit and consistency across studies	*Guideline #3, Recommendation A*: "Diagnostic testing is most cost-effective when it leads to a change in treatment that improves outcomes, not just a change in treatment. Although diffusion-weighted magnetic resonance imaging (DW-MRI) is more sensitive than CT for detecting AIS, routine use in all patients with AIS is not cost-effective."
What harms were identified?	The guidelines do not provide details about potential harms associated with expedited brain imaging of acute stroke patients that were identified in the body of evidence.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	Additional evidence identified as part of the contractor's annual review of the clinical literature is described in Section **1a.4**.

1a.4 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.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

Additional evidence supporting the measure was identified through a review of clinical literature and related policy, as further described and referenced in Sections **1a.4.2** and **1a.4.3**. These additional references provide evidence that stroke remains the fifth leading cause of death and that improved outcomes, based on practice and studies, is associated with early interventions and timeframes from door-to-imaging and door-to-treatment for patients presenting with acute ischemic stroke.

1a.4.2 What process was used to identify the evidence?

In addition to the three 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 and October 1, 2017 to March 31, 2019. Search results were limited to those published in the English language and that had abstracts available in PubMed. A further review by the contractor's clinical and measure-development team resulted in the inclusion of eight articles in the body of evidence below. Citations and summaries for the eight items included in this review can be found in Section **1a.4.3**.

1a.4.3. Provide the citation(s) for the evidence.

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):904–913.

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.

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):225–243.

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.

Haršány M, Tsivgoulis G, Alexandrov A. Intravenous thrombolysis in acute ischemic stroke: Standard and potential future applications. *Expert Review of Neurotherapeutics*. 2014; 14(8):879–892.

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.

Heron, Melonie. Deaths: Leading causes for 2016. *National Vital Statistics Reports*. 2018; 67(6). https://www. cdc.gov/ nchs/data/nvsr/nvsr67/nvsr67_06.pdf.

This report shares the leading causes of death in the United States by age, sex, race, and Hispanic origin using 2016 data from 50 states and the District of Columbia.

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):2914–2919.

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.

Metts, E. L., A. M. Bailey, K. A. Weant and S. B. Justice. Identification of rate-limiting steps in the provision of thrombolytics for acute ischemic stroke. *J Pharm Pract.* 2017; 30(6), 606–611.

A retrospective chart review identified a number of factors that contribute to delays in DTN times greater than the guideline recommended 60 minutes. The study suggested that patient care protocols should focus on reducing *potential* delays in stroke treatment.

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):543–549.

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.

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., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

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).

REFERENCES:

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) 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.

1b.2. Provide performance scores on the measure as specified (<u>current and over time</u>) at the specified level of analysis. *(<u>This is required for maintenance of endorsement</u>. 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 include.) This information also will be used to address the sub-criterion on improvement (4b1) 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 July 2017 through June 2018 data collection reporting period, performance scores ranged from 0.0% to 100%, with a weighted mean of 75.0, representing an 25.8% increase in the weighted mean of performance scores from the October 2012 through September 2013 data collection period.

The data presented below represent performance scores for the facilities whose denominator counts met minimum case count requirements during the October 2012 through September 2013 and July 2017 through June 2018 data collection periods. 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 12 months of data available (July 2017 through June 2018) were compared against the October 2012 through September 2013 data.

Further details on the descriptive statistics for longitudinal facility performance are included below:

	Data Collection Period	Percentage Point Change1
	October 2012–September 2013	July 2017–June 2018
Facilities	918	1,550

Minimum Value	0.0%	0.0%	0.0
1st Percentile	4.0%	17.0%	13.0
5th Percentile	15.0%	36.0%	21.0
10th Percentile	24.0%	46.0%	22.0
25th Percentile	42.0%	64.0%	22.0
Median	62.0%	79.0%	17.0
75th Percentile	78.0%	88.0%	10.0
90th Percentile	88.0%	94.0%	6.0
95th Percentile	92.0%	100.0%	8.0
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)	75.0% (82.2)
Number of CT and MRI scans performed (Denominator)		16,817	31,939

(1) Note that this value represents the percentage point change, not the percentage increase/decrease.

From the inception of public reporting through June 2018 data collection, there has been wide variation in facility performance, though this variation has narrowed. During the October 2012 through September 2013 data collection period, the interquartile range (IQR) of performance scores ranged from 42% to 78%. During the July 2017 through June 2018 reporting period, the IQR ranged from 64% to 88%. 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, comparing data from October 2012 through September 2013 to data from July 2017 through June 2018.

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 maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) 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. The same analysis was subsequently repeated using CDW data from July 2016 through December 31, 2018, with 80,749 patients meeting the measure's criteria for numerator inclusion.

In 2014, 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).

In contrast, the July 2016 through December 31, 2018 data showed no race disparities for head CT or MRI scan interpreted within 45 minutes of ED arrival; however, Hispanic patients remained less likely to be included in the measure's numerator (OR=0.81, p=0.001) compared to non-Hispanic patients. Females also remained less likely than males to have a head CT or MRI scan interpreted within 45 minutes of ED arrival (OR=0.87, p<0.001).

Facility characteristics also play 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. In 2014, 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).

Logistic regression performed on data from July 2016 through December 31, 2018 showed a similar pattern, with patients treated in larger facilities, that is, facilities with 51-100 beds (OR=1.18, p=0.002), 101 to 250 beds (OR=1.35, p<0.001), 251-500 beds (OR=1.35, p<0.001) significantly more likely to have a head CT or MRI scan interpreted within 45 minutes of ED arrival than patients in facilities with 50 or fewer beds. Similar to 2014, patients treated in major teaching hospitals 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=0.77, p=0.013).

While the current and previous logistic regression models 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 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

Data on disparities from the measure as specified is reported in 1b.4, above.

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 sub criteria 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. Non-Condition Specific*(check all the areas that apply):*

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

Elderly, Populations at Risk

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/inpatient/specifications-manuals#tab2

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:** AppendixA_v12.0a_010119_0930190.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Yes

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update 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.

Since endorsement in 2016, ICD-10-CM diagnosis codes have been updated and provided as a part of the annual measure maintenance process. The most recent annual update review occurred in October 2018.

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) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Emergency department (ED) acute ischemic stroke or hemorrhagic stroke patients arriving at the ED within 2 hours of the time last known well, with an order for a head CT or MRI scan whose time from ED arrival to interpretation of the Head CT scan is within 45 minutes of arrival.

S.5. 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, time period for data collection, 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 (S.14).

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

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

Emergency department acute ischemic stroke or hemorrhagic stroke patients arriving at the ED within two hours of the time last known well with an order for a head CT or MRI scan.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, 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 target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

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.8. 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.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, 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)

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – 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.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

S.13. 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.14. Calculation Algorithm/Measure Logic (*Diagram or 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; time period for data, 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

Measure = numerator counts / denominator counts [The value should be recorded as a percentage]

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., 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.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

This measure does not use survey data.

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Claims, Electronic Health Records, Paper Medical Records

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(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.19. 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.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Facility, Other

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Emergency Department and Services

If other:

S.22. <u>COMPOSITE Performance Measure</u> - 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.

2. Validity – See attached Measure Testing Submission Form

NQF_0661_Measure_Testing_Form.docx

2.1 <u>For maintenance of endorsement</u>

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.2 <u>For maintenance of endorsement</u>

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.3 <u>For maintenance of endorsement</u>

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

No - This measure is not risk-adjusted

Measure Testing (subcriteria 2a2, 2b1-2b6)

1. DATA/SAMPLE USED FOR <u>ALL</u> 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. <u>If there are differences by aspect of testing</u>,(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.17*)**	**Measure Tested with Data From:**
☒ ☒ abstracted from paper record	☒ ☒ abstracted from paper record
☒ ☒ claims	☒ ☒ claims
☐ ☐ registry	☐ ☐ 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. 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
  - 99283
  - 99284
  - 99285
  - 99291
- The initial patient population includes cases for patients 18 years and older, as of the date of the encounter, with a principle diagnosis for acute ischemic or hemorrhagic stroke, identified by using any of the following International Classification of Diseases, version 10 (ICD-10) codes
  - I6000
  - I6001
  - I6002
  - I6010
  - I6011
  - I6012
  - I602
  - I6030
  - I6031
  - I6032
  - I604
  - I6050
  - I6051
  - I6052
  - I606
  - I607
  - I608
  - I609
  - I610
  - I611
  - I612
  - I613
  - I614
  - I615
  - I616
  - I618
  - I619
  - I629
  - I6300
  - I63011
  - I63012
  - I63013
  - I63019
  - I6302
  - I63031
  - I63032
  - I63033
  - I63039
  - I6309
  - I6310
  - I63111
  - I63112
  - I63113
  - I63119
  - I6312
  - I63131
  - I63132
  - I63133
  - I63139
  - I6319
  - I6320
  - I63211
  - I63212
  - I63213
  - I63219
  - I6322
  - I63231

- I63232
- I63233
- I63239
- I6329
- I6330
- I63311
- I63312
- I63313
- I63319
- I63321
- I63322
- I63323
- I63329
- I63331
- I63332
- I63333
- I63339
- I63341
- I63342
- I63343
- I63349
- I6339
- I6340
- I63411
- I63412
- I63413
- I63419
- I63421
- I63422
- I63423
- I63429
- I63431
- I63432
- I63433
- I63439
- I63441
- I63442
- I63443
- I63449
- I6349
- I6350
- I63511
- I63512
- I63513
- I63519
- I63521
- I63522
- I63523
- I63529
- I63531
- I63532
- I63533
- I63539
- I63541
- I63542
- I63543
- I63549
- I6359
- I636
- I6381
- I6389
- I639

2. 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, who arrived to the emergency department (ED) within 2 hours of the *Time Last Known Well* and had an order for a head CT or MRI scan.
3. 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"
     - Difference between *Arrival Time* and *Time Last Known Well* is greater than 120 minutes

4. 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
5. 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, 2014 | July 1, 2016–December 31, 2018

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.20*)**	**Measure Tested at Level of:**
☐ ☐ individual clinician	☐ ☐ individual clinician
☐ ☐ group/practice	☐ ☐ group/practice
☒ ☒ hospital/facility/agency	☒ ☒ hospital/facility/agency
☐ ☐ health plan	☐ ☐ health plan
☒ ☒ other: National | National	☒ ☒ other: national | National

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.

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.

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

Reliability Testing

Data Source: Hospital Compare downloadable file [maintained by the Centers for Medicare & Medicaid Services (CMS)]

Dates: *Denominator:* July 1, 2017–June 30, 2018; *Numerator:* July 1, 2017–June 30, 2018; *Exclusions:* July 1, 2017–June 30, 2018; *Exceptions:* July 1, 2017–June 30, 2018

Number of Facilities: 1,550

Denominator Cases: 31,939

Numerator Cases: 23,953

Level of Analysis: Facility

Patient Characteristics: Not applicable

Validity Testing—*Empirical Validity*

Data Source: Clinical Data Abstraction Center (CDAC), Clinical Data Warehouse (CDW)

Dates: July 1, 2016–September 30, 2018

Sampled Population: 2,622

Level of Analysis: Data element

Patient Characteristics: Not applicable

Exclusions Analysis

Data Source: Denominator: CDW; Numerator: CDW; Exclusions: CDW

Dates: *Denominator:* July 1, 2016—December 31, 2018; *Numerator:* July 1, 2016—December 31, 2018; *Exclusions:* July 1, 2016—December 31, 2018; *Exceptions:* July 1, 2016—December 31, 2018

Number of Facilities: 3,140

Sampled Population: 271,756

Denominator Cases: 271,756

Numerator Cases: 80,749

Level of Analysis: Case

Denominator Patient Characteristics: Gender (% Male): 49.7; Mean Age (Years): 67.1 (St. Dev.: 15.3); Race (% Minority): 16.4

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:* July 1, 2017–June 30, 2018; *Numerator:* July 1, 2017–June 30, 2018; *Exclusions:* July 1, 2017–June 30, 2018; *Exceptions:* July 1, 2017–June 30, 2018

Number of Facilities: 1,550

Denominator Cases: 31,939

Numerator Cases: 23,953

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: Hospital Compare downloadable file

Dates: *Denominator:* July 1, 2017–June 30, 2018; *Numerator:* July 1, 2017–June 30, 2018; *Exclusions:* July 1, 2017–June 30, 2018; *Exceptions:* July 1, 2017–June 30, 2018

Number of Facilities: 1,550

Denominator Cases: 31,939

Numerator Cases: 23,953

Level of Analysis: Facility

Patient Characteristics: Not applicable

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

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**.

We assessed patient-level sociodemographic status (SDS) factors as part of the regression model reported in Section **2b3.2**, 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 **2b3.2.**

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.

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 July 1, 2017–June 30, 2018 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.

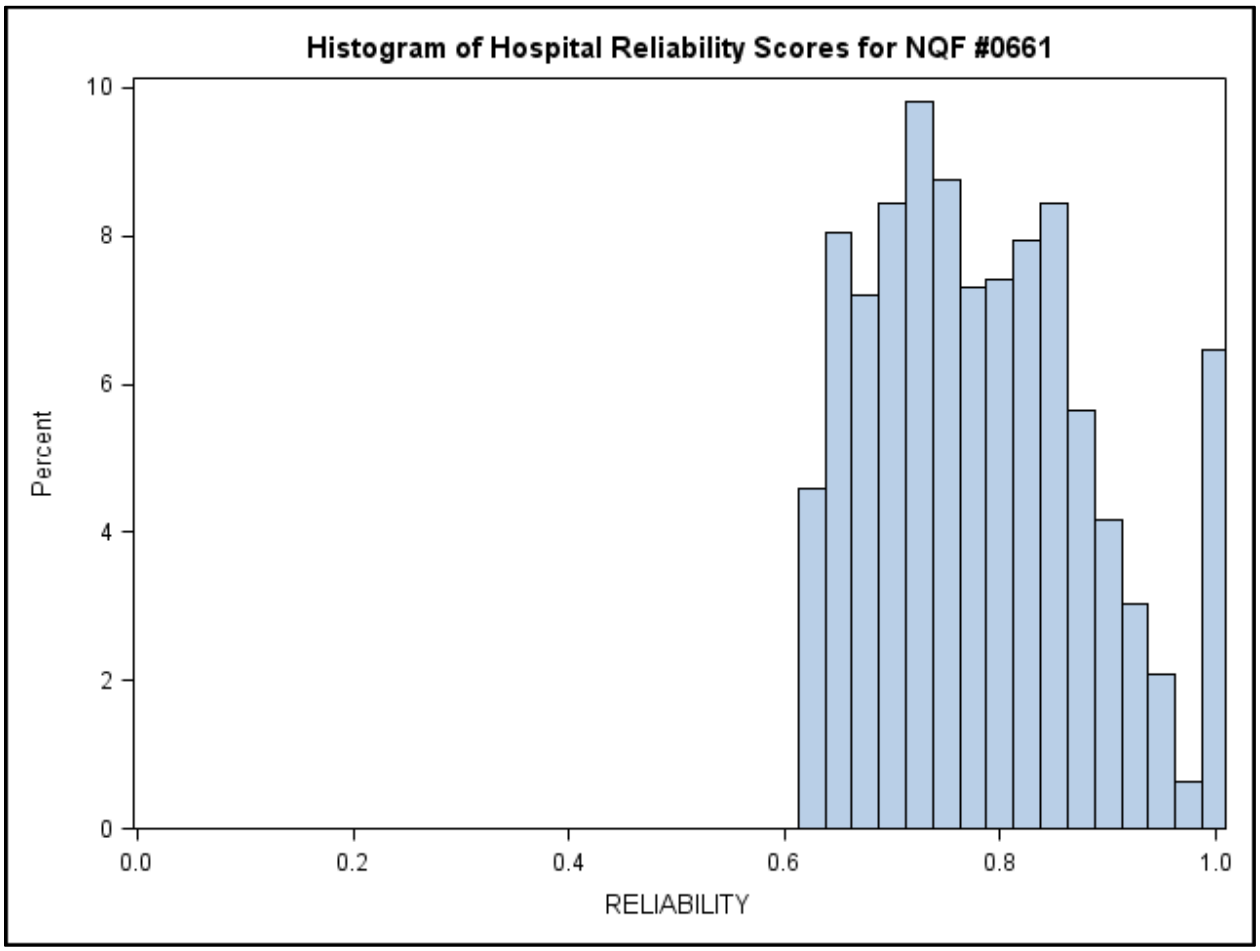
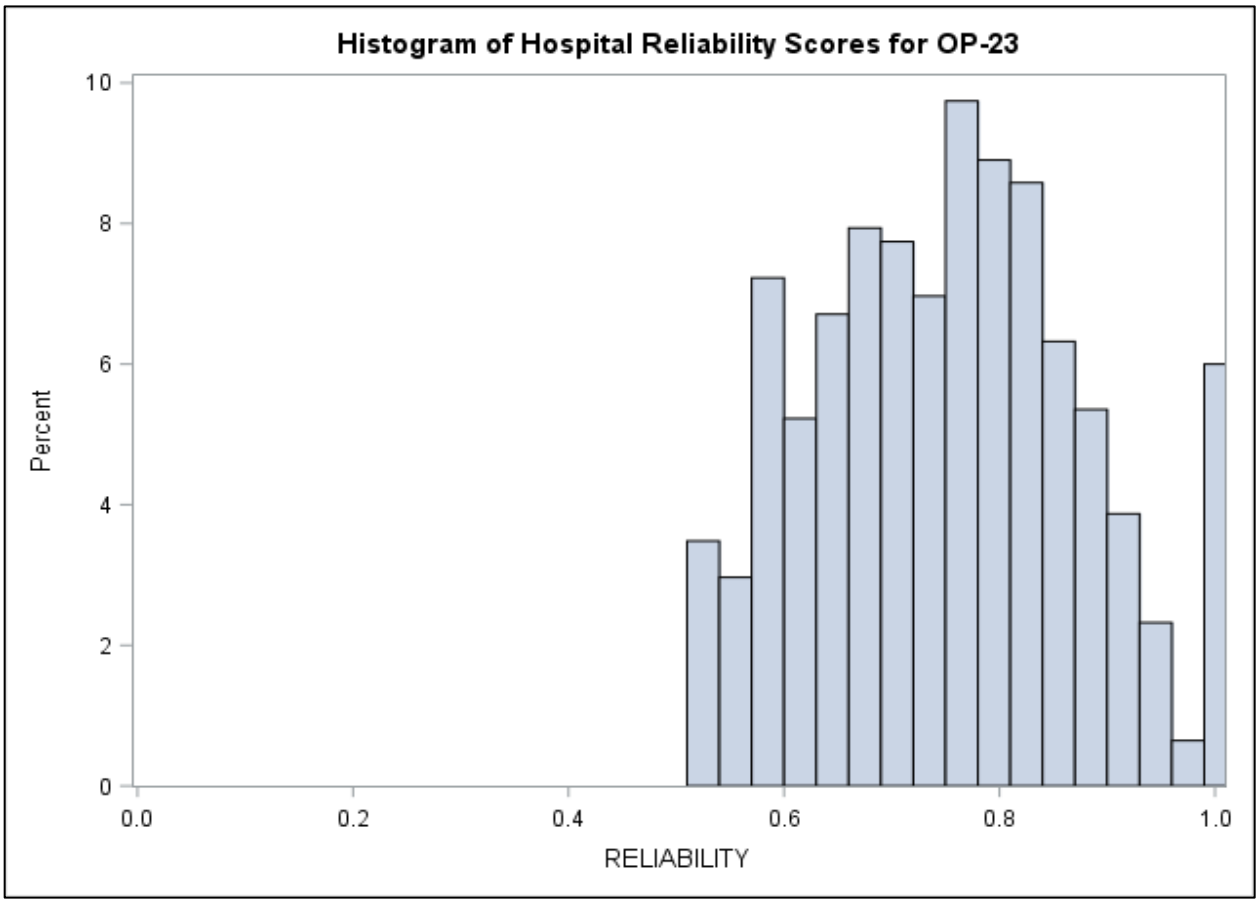


Figure 1 (below) is a histogram of the distribution of the reliability scores for the facilities meeting the minimum case count requirements during the July 1, 2017 – June 30, 2018 data collection period. Reliability scores ranged from 0.52 to 1.00, with a median reliability score of 0.76. Higher scores 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.

Calculated using a beta-binomial model, a median reliability score of 0.76 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.

2b1. VALIDITY TESTING

2b1.1. 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*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. 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)*

We assessed the validity of the measure using both qualitative and quantitative analyses. Empirical validity of the data elements was assessed by calculating a rate of agreement between facility abstraction and auditor (CDAC) abstraction for each of the data elements used to calculate the measure. 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 to March 31, 2015.

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.

Empirical validity of critical data elements was assessed by examining kappa statistics (for categorical variables and the constructed outcomes of the numerator and denominator) and Pearson's correlation coefficient (for non-categorical variables) between facility abstraction and auditor (CDAC) abstraction for each of the data elements used to calculate the measure. This analysis used data elements for 2,622 cases abstracted by CDAC, which were previously abstracted by facilities; these data were collected from July 1, 2016 to December 31, 2018.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Results of empirical validity testing 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%.

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

Results of the data element validity analysis are presented and organized by variable type in the tables below.

Table 1. Data Element Validity for Categorical Variables

Variable (Categorical)	Kappa Value	p Value	n
Discharge	0.9318	0.0000	2,617
Head CT/MRI Scan	0.8625	0.0000	2,471
Last Known Well	0.7676	0.0000	2,394

Table 2. Data Element Validity for Non-Categorical Variables

Variable (Non-Categorical)	Pearson's Correlation Coefficient	p Value	n
Date Last Known Well (LKW)	1.0000	p < .001	1,391
Date of CT/MRI Scan	1.0000	p < .001	668
Time of Arrival	0.5122	p < .001	95
Time LKW	0.9288	p < .001	1,386
Time of CT/MRI Scan	0.9882	p < .001	659

Table 3. Validity of Constructed Outcomes

Variable (Categorical)	Kappa Value	p Value	n
Numerator	0.8450	0.0000	2,622
Denominator	0.8464	0.0000	2,622

2b1.4. 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. The rate of agreement between facility and CDAC abstraction ranged from moderate to strong across the data elements used to calculate OP-23. The rate of agreement was strong for dichotomous variables, as well as those based on administrative data. Agreement was moderate for clinical variables related to time.

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 agreement between facility and CDAC-abstracted data elements ranged from moderate to strong across the data elements used to calculate OP-23. Kappa values ranged from 0.77–0.93 for categorical data elements; Pearson's correlation coefficients for non-categorical variables ranged from 0.51-0.92; and Kappa values for the constructed variables of the numerator and denominator were each 0.85. These results support the validity of the measure and its calculation.

2b2. EXCLUSIONS ANALYSIS

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

2b2.1. 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.

We tested measure exclusions and numerator exceptions to determine the prevalence of each exclusion and exception, by facility, and overall. The analysis tested measure exclusions and numerator exceptions during the July 1, 2016—December 31, 2018 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 **1.2e**, above.

[1] Respondents were provided a table detailing the key measure exclusions.

2b2.2. 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
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	-	-	-

Data Element	Denominator Exclusion or Numerator Exception?		Overall Occurrence		Distribution Across Facilities (%)		
	Denominator Exclusion	Numerator Exception	N	%	25th	50th	75th
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.

We examined overall frequencies and proportions of cases excluded for each exclusion/exception criterion, among all sampled cases, for 3,140 facilities submitting eligible cases from July 1, 2016—December 31, 2018. The sampled population included 271,756 cases where a patient (ages 18 years or older) presented with an ischemic or hemorrhagic stroke to an ED.

Table 4. 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		10,702	3.94	2.0	4.0	7.0
Head CT Scan or MRI Order	X		7,635	2.92	1.0	2.0	5.0
Last Known Well	X		101,127	37.21	27.0	44.0	69.0
Date Last Known Well		X	0	0.00	0.0	0.0	0.0
Time Last Known Well		X	192	0.00	0.0	0.0	0.0
Arrival Time		X	0	0.0	0.0	0.0	0.0
Last Known Well Minutes	X		77,757	39.38	22.0	34.0	52.0

Data Element	Denominator Exclusion or Numerator Exception?		Overall Occurrence		Distribution Across Facilities (%)		
	Denominator Exclusion	Numerator Exception	N	%	25th	50th	75th
Head CT Scan or MRI Interpretation Date		X	0	0.00	0.0	0.0	0.0
Head CT Scan or MRI Interpretation Time		X	54	0.00	0.0	0.0	0.0
Total Denominator Exclusions	4 exclusions	-	197,221	72.66	-	-	-
Total Numerator Exceptions	-	5 exceptions	246	0.00	-	-	-
Total Removed from the Denominator or Numerator	9 exceptions and exclusions		197,467	72.66	-	-	-

As indicated in the table above, 72.7% of the initial patient population is excluded from the denominator. 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.

2b2.3. 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.

As seen in table **2b2.2** above, a large number of patients (72.7%) are excluded from the measure calculation; the removal of cases where an abstractor submits a value of "UTD" for *Last Known Well* is necessary to align

with clinical guidelines and enable measure calculation. Missing data on this criterion makes it impossible to determine the timeliness of interpretation of a head CT or MRI. 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, 3.94% 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 interquartile range of 2.0% to 7.0%.
- *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 1.0 to 5.0%.
- *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.* Patients where *Last Known Well* is equal to "No" are excluded from the measure denominator. Overall, 37.21% 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 27.0% to 69.0%.
- *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. No cases of patients included in the denominator had "UTD" value for *Date Last Known Well*. 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, <0.1% 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. No patients included in the denominator have a "UTD" value for *Arrival Time*. 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, 28.61% 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 22.0% to 52.0%.
- *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. No patients included in the denominator have a "UTD" value for *Head CT Scan or*

MRI Interpretation Date. 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, <0.1% of patients included in the denominator have a "UTD" value for *Head CT Scan or MRI Interpretation Date*. 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.

2b3. 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.

2b3.1. 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

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

Not applicable.

Not applicable.

2b3.2. If an outcome or resource use component 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.

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.

Results of logistic regression of inclusion in the measure's numerator (i.e., having timely order and interpretation of head CT or MRI) on patient characteristics are provided in *Table 5*, below. Odds of inclusion in the measure's numerator are significantly lower for beneficiaries in younger age ranges as compared to the referent group (60

to 69); whereas, odds of inclusion are marginally higher (for 70 to 79) or significantly higher (for 80 to 89 and 90 and older) for older age ranges. Compared to males, females were less likely to be included in the numerator. Race was not associated with odds of inclusion in the numerator; however, Hispanic/Latino ethnicity was associated with lower odds of inclusion in the numerator. Compared with the smallest facilities (i.e., those with 50 or fewer beds) beneficiaries at larger facilities up to 500 beds had higher odds of numerator inclusion. Facility location was not associated with odds of inclusion in the numerator. Compared to non-teaching facilities, beneficiaries at major teaching facilities had significantly lower odds of inclusion in the numerator.

Table 5. Logistic Regression Results

Variable	*Odds Ratio*	*CI Lower Bound*	*CI Upper Bound*	*P Value*
Beneficiary Age				
18 to 29	0.460	0.400	0.530	0.000
30 to 39	0.612	0.555	0.674	0.000
40 to 49	0.773	0.724	0.825	0.000
50 to 59	0.942	0.892	0.995	0.032
60 to 69 (Referent)	--	--	--	--
70 to 79	1.051	0.999	1.106	0.057
80 to 89	1.192	1.127	1.261	0.000
90 and Older	1.196	1.097	1.304	0.000
Beneficiary Gender				
Male (Referent)	--	--	--	--
Female	0.874	0.845	0.905	0.000
Unknown	1.661	0.388	7.113	0.494
Beneficiary Race				
Caucasian (Referent)	--	--	--	--
African-American	0.942	0.878	1.010	0.095
Asian	0.985	0.844	1.150	0.850
America Indian or Alaska Native	0.813	0.653	1.012	0.064
Native Hawaiian or Pacific Islander	0.826	0.590	1.156	0.265
Unknown Racial or Ethnic Identification	0.917	0.795	1.057	0.231
Ethnicity				
Non-Hispanic/Latino (referent)	--	--	--	--
Hispanic/Latino	0.805	0.707	0.915	0.001
Facility Bed Count				
0 to 50 Beds (Referent)	--	--	--	--
51 to 100 Beds	1.187	1.066	1.322	0.002
101 to 250 Beds	1.353	1.221	1.499	0.000
251 to 500 Beds	1.348	1.195	1.521	0.000
500+ Beds	0.876	0.744	1.030	0.109
Unknown Bed Count	-	-	-	-

Variable	*Odds Ratio*	*CI Lower Bound*	*CI Upper Bound*	*P Value*
Facility Location				
Rural (Referent)	--	--	--	--
Urban	0.991	0.912	1.076	0.821
Location Unknown	1.654	1.113	2.456	0.013
Facility Teaching Status				
Non-Teaching (Referent)	--	--	--	--
Teaching	1.046	0.936	1.168	0.426
Major Teaching	0.770	0.627	0.946	0.013
Unknown Teaching Status	0.394	0.344	0.452	0.000
Constant	2.958	2.712	3.226	0.000

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.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk 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*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

Not applicable—No risk adjustment or stratification was performed.

Not applicable—No risk adjustment or stratification was performed.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☒ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable—No risk adjustment or stratification was performed.

Not applicable—No risk adjustment or stratification was performed.

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk 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.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

Not applicable—No risk adjustment or stratification was performed.

Not applicable—No risk adjustment or stratification was performed.

2b3.5. 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—No risk adjustment or stratification was performed.

Not applicable—No risk adjustment or stratification was performed.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below. ***If stratified, skip to 2b3.9***

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable—No risk adjustment or stratification was performed.

Not applicable—No risk adjustment or stratification was performed.

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable—No risk adjustment or stratification was performed.

Not applicable—No risk adjustment or stratification was performed.

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable—No risk adjustment or stratification was performed.

Not applicable—No risk adjustment or stratification was performed.

2b3.9. Results of Risk Stratification Analysis:

Not applicable—No risk adjustment or stratification was performed.

Not applicable—No risk adjustment or stratification was performed.

2b3.10. 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.

Not applicable—No risk adjustment or stratification was performed.

2b3.11. 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.

Not applicable—No risk adjustment or stratification was performed.

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. 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**.

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 July 1, 2017–June 30, 2018 data, this included 1550 facilities. For each facility, the facility performance score and standard deviation was

calculated. This analysis identified 92 facilities as statistical outliers. Additional details of this analysis are provided in Section **2b4.2**.

2b4.2. 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.

Of the 1,550 facilities reporting during the July 1, 2017–June 30, 2018 data collection period, 92 (5.9%) 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. *Table 6*, below, displays the distribution of performance scores for OP-23 across common percentiles. The distribution of performance scores, combined with the identification of outlier facilities, demonstrates the ability of the measure to meaningfully distinguish differences in facility performance.

Table 6. Distribution of Performance Scores (among Facilities Meeting Minimum Case Count)

	Mean (SD)	*1%*	*5%*	*10%*	*25%*	*50%*	*75%*	*90%*	*95%*	*99%*
Performance	77.11% (14.84)	28.57	48.48	57.14	69.57	80.00	87.50	92.94	95.65	100.00

2b4.3. 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.

Analysis of the July 1, 2017–June 30, 2018 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.

2b5. 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 social risk 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 social risk***

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.

2b5.1. 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—No risk adjustment or stratification was performed.

Not applicable—No risk adjustment or stratification was performed.

2b5.2. 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—No risk adjustment or stratification was performed.

Not applicable—No risk adjustment or stratification was performed.

2b5.3. 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—No risk adjustment or stratification was performed.

Not applicable—No risk adjustment or stratification was performed.

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. 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.**

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 **2b2.3.**

2b6.2. 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.

The frequency and distribution of missing data are described in Section **2b2.3**. We did not perform statistical analyses of missing data.

2b6.3. 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.

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.

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*)
Update this field for **maintenance of endorsement**.

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. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

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. Please also complete and attach the NQF Feasibility Score Card.

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. Required for maintenance of endorsement. Describe difficulties (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 instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

NQF 0661 was added to the Hospital OQR Program in November 2014 via the Outpatient Prospective Payment System (OPPS) rule for calendar year 2015 (79 FR 66769). Since that time, CMS has not received feedback from stakeholders suggesting difficulty with data collection for the measure, nor its feasibility or usability.

In fall 2015, 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. Unstructured feedback from the EWG obtained in summer 2019 is consistent with the 2015 survey.

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 2015 American Medical Association. All rights reserved. CPT is a registered

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.

Specific Plan for Use	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 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 (external benchmarking to organizations) Hospital Outpatient Quality Reporting http://www.medicare.gov/hospitalcompare/search.html Hospital Outpatient Quality Reporting Hospital Outpatient Quality Reporting https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1192804531207

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

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

Public Reporting:
Name of program and sponsor: The CMS Hospital OQR Program
Purpose: The Hospital OQR 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 Hospital OQR 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 Hospital OQR 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 participating in the Hospital OQR Program in the United States that meet minimum case count requirements. During the July 2017 through June 2018 data collection period, 1550 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 Hospital OQR Program
Purpose: The Hospital OQR 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.

4a1.2. 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.

4a1.3. 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.

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

As a part of the Hospital OQR Program, performance data is provided quarterly to facilities in their preview reports, which are distributed to hospitals via their QualityNet secure inboxes, and are available to the public through the CMS Hospital Compare website. The preview report contains a summary of the facility's score on a number of metrics, including NQF 0661. All facilities eligible to report NQF 0661 receive a preview report and FSR each quarter, regardless of whether they have sufficient cases to publicly report NQF 0661. Technical assistance is provided through help desk support (RightNow Q&A Tool) and Hospital OQR Program webinars. Frequently asked questions are monitored and used to determine whether updates or additional guidance is needed for facilities.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

As a part of the Hospital OQR Program, performance data is provided quarterly to facilities in their preview reports, which are distributed to hospitals via their QualityNet secure inboxes, and are available to the public through the CMS Hospital Compare website. The preview report contains a summary of the facility's score on a

number of metrics, including NQF 0661. All facilities eligible to report NQF 0661 receive a preview report and FSR each quarter, regardless of whether they have sufficient cases to publicly report NQF 0661. On Hospital Compare the performance data is located under Timely and Effective Care – ED Throughput. To assist users with understanding the data on Hospital Compare, CMS shares details on the data collection period of what is currently displayed (10/1/2017–9/30/2018) and user-friendly language describing the measure title ("Percentage of patients who came to the emergency department with stroke symptoms who received brain scan results within 45 minutes of arrival").

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

Feedback received from stakeholders (via the RightNow Q&A tool) is used to revise the measure specifications. Following receipt of a suggestion to adjust the specifications, a literature review is performed to determine if the proposed change aligns with the empirical evidence base for the measure; feedback from the expert work group is obtained to evaluate the change to the specifications. To date, we have received no significant concerns raised by stakeholders about the measure specifications through the RightNow Q&A tool.

In addition, stakeholders may submit comments on the measure through the Outpatient Prospective Payment System (OPPS) annual rule-making process. No comments were received for this measure during the Calendar Year (CY) 2016 -2019 OPPS rule-making cycles.

4a2.2.2. Summarize the feedback obtained from those being measured.

To date, we have received no significant feedback about the measure specifications.

4a2.2.3. Summarize the feedback obtained from other users

To date, we have received no significant feedback about the measure specifications.

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

To date, we have received no significant feedback about the measure specifications.

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.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (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.)

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.

Summary statistics of performance scores during the October 2012 through September 2013 and July 2017 through June 2018 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 27percentage points (from 62.0% to 79.0%) between 2012 and 2018. Nine hundred eighteen facilities met minimum case counts during the October 2012 through September 2013 data collection periods, and 1550 facilities met minimum case counts during the July 2017 through June 2018 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 July 2017 through June 2018 data collection period, 31,939 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, 23,955 had a head CT or MRI scan interpreted within 45 minutes of ED arrival (75.0%).

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.

4b2. 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).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

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.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

We did not identify any unintended consequences from implementing this 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.

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 of Related Measures

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 harmonized to the extent possible?

Yes

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 (Hospital OQR), the two measures serve different target populations and purposes: the Hospital OQR 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 Hospital OQR 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. While the ICSI measure is related to NQF #0661, it focuses on head CT completion, which is an intermediate step for head CT interpretation (NQF #0661). NQF #0661 includes an additional imaging modality—MRI interpretation. Details about the measure algorithm, data elements, and measure specifications for the ICSI measure are not readily available to compare.

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 and Medicaid Services

Co.2 Point of Contact: Joseph, Clift, Joseph.Clift@cms.hhs.gov, 410-786-4165-

Co.3 Measure Developer if different from Measure Steward: Mathematica

Co.4 Point of Contact: Robert, Dickerson, RDickerson@mathematica-mpr.com, 312-585-3345-

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 expert workgroup (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:

Kenneth Bricker, DO

EWG 2019; Minneapolis VA Medical Center

Joseph P. Drozda, Jr., MD

TEP 2010; Mercy Health, Rep. of American College of Cardiology; Director of Outcomes Research

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

EWG 2019; Emergency Nurses Association (ENA), Institute for Quality, Safety, and Injury Prevention, Director

David Seidenwurm, MD

EWG 2019; 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

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, 2019

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? 08, 2019

Ad.6 Copyright statement: This measure does not have a copyright.

Ad.7 Disclaimers: CPT codes, descriptions, and other data only are copyright 2015 American Medical Association. All rights reserved. CPT is a registered trademark of the American Medical Association (AMA). Applicable Federal Acquisition Regulation Site (FARS)\Defense Federal Acquisition Regulation Statement

(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.

This performance measure is not a clinical guideline and does not establish a standard of medical care, and has not been tested for all potential applications. The measure and specifications are provided without warranty.

Ad.8 Additional Information/Comments: