NATIONAL QUALITY FORUM
Driving measurable health improvements together

Memo

November 30, 2021

To: Consensus Standards Approval Committee (CSAC)

From: Neurology Project Team

Re: Neurology Spring 2021 Cycle

CSAC Action Required

The CSAC will review recommendations from the Neurology project at its November 30 and December 1, 2021, meeting and vote on whether to uphold the recommendations from the Committee.

This memo includes a summary of the project, measure recommendations, themes identified, responses to the public and member comments, and the results from NQF member expression of support. The following document accompany this memo:

1. **Neurology Spring 2021 Draft Report**. The draft report has been updated to reflect the changes made following the Standing Committee's discussion of public and member comments. The complete draft report and supplemental materials are available on the project webpage.

Background

Neurological conditions are disorders that affect the brain and the nerves found throughout the body and spinal cord. NQF's Neurology Standing Committee assesses new and existing measures related to brain and spinal conditions brought by measure developers for endorsement. For the spring 2021 cycle, the NQF Neurology project evaluated one new measure and one maintenance measure. The new measure assessed the rate of missed stroke in emergency departments (EDs) when patients present themselves with dizziness, which is a nonspecific symptom. The maintenance measure correlated to the appropriate measurement of carotid stenosis (i.e., a narrowing of the carotid artery) by radiologists on imaging studies, including computed tomography (CT), angiography, ultrasound, and magnetic resonance imaging (MRI).

The Standing Committee recommended the following measure:

- **#0507** Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports (American College of Radiology) (maintenance)

The following measure was withdrawn from consideration following the post-comment meeting:

- **#3614** Hospitalization After Release With Missed Dizzy Stroke (Johns Hopkins Armstrong Institute of Patient Safety and Quality) (new)

Draft Report

The Neurology Spring 2021 draft report presents the results of the evaluation of two measures considered under the Consensus Development Process (CDP). One is recommended for endorsement. One measure was withdrawn after the post-comment meeting.

The measures were evaluated against the 2019 version of the measure evaluation criteria.

Measures under Review	Maintenance	New	Total
Measures under review	1	1	2
Measures recommended for endorsement	1	0	1
Measures withdrawn from consideration	0	1	1
Reasons for not recommending	Importance – 0 Scientific Acceptability -0 Use - 0 Overall - 0 Competing Measure -0	Importance - 0 Scientific Acceptability - 0 Use - 0 Overall - 0 Competing Measure – 0	0

CSAC Action Required

Pursuant to the CDP, the CSAC is asked to consider endorsement of one candidate measure and to uphold the standing committee's recommendation to not endorse one measure.

Measures Recommended for Endorsement

- **#0507** Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports (American College of Radiology) maintenance

  Overall Suitability for Endorsement: Yes-12; No-0 (denominator = 12)

Comments and Their Disposition

NQF received five comments from five organizations (including five member organizations) and individuals pertaining to the draft report and to the measures under review.

A narrative of all comments submitted during the comment period, with the responses to each comment and the actions taken by the Standing Committee and measure developers, is posted to the Neurology project webpage.

Comment Themes and Committee Responses

Comments about specific measure specifications and rationale were forwarded to the developers, who were invited to respond.

The Standing Committee reviewed all of the submitted comments (general and measure specific) and developer responses. Committee members focused their discussion on measures or topic areas with the most significant and recurring issues.

Measure-Specific Comments

#0507: Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports

The American College of Radiology (ACR) has completed data element empirical validity to support NQF 0507 re-endorsement. According to the Blueprint for the Centers for Medicare & Medicaid Services (CMS) Measures Management System, data element validity is the "extent to which the information represented by the data element or code used in the measure reflects the actual concept or event intended." The ACR performed random audits using the groups that submitted Qualified Clinical Data Registry (QCDR) records for the measure to CMS for their Merit-based Incentive Payment System (MIPS) program, over a four-year period. The auditors compared the numerator data element registry submissions used in measure calculation with actual exam records from the submitters' systems. The audit confirmed a high level of agreement and concordance between the data shown on exam records and what was submitted to the registry. The records where exam data did not match the registry data represent human error in collection or submission of data.

Summary of NQF 0507 in 2017-2020 Audit by Year

Year	# Groups	# Records Audited	# Records Without Issue	% Records Without Issue
2017	11	108	106	98%
2018	22	128	126	98%
2019	17	130	130	100%
2020	15	69	69	100%

Measure Steward/Developer Response:

N/A

Committee Response:

Thank you for your comment. The Standing Committee reviewed and considered this information during the post-comment meeting and found the measure's validity acceptable and the measure to be suitable for endorsement at this time.

Member Expression of Support

Throughout the 16-week continuous public commenting period, NQF members had the opportunity to express their support ('support' or 'do not support') for each measure submitted for endorsement consideration to inform the Committee's recommendations. One NQF member provided their expression of non-support for measure 3614, which was withdrawn from consideration following the post-comment meeting. Appendix C details the expression of support.

Appendix A: CSAC Checklist

The table below lists the key considerations to inform the CSAC's review of the measures submitted for endorsement consideration.

Key Consideration	Yes/No	Notes
Were there any process concerns raised during the CDP project? If so, briefly explain.	Yes	The developer of the withdrawn measure (#3614) raising concerns that the initial measure evaluation meeting discussion was fragmented due to loss of quorum and loss of sufficient attendance to hold the meeting, at different times. The developer also objected to one of the measure developers being unable to act as the main presenter during the meeting since he was also a Standing Committee member who was recused from discussion and voting on that measure. NQF maintains that attendance, discussion, and voting policies were followed consistently throughout the process. In addition, although one of the developers was unable to act as the main presenter during the meeting, he was permitted to respond to Standing Committee questions if no one else on his team was able to answer them.
Did the Standing Committee receive requests for reconsideration? If so, briefly explain.	Yes	The Standing Committee received a reconsideration request from the developer of #3614, which failed the evidence criterion during the measure evaluation meetings. The measure again did not pass the Evidence criterion during the post-comment meeting and the developer decided to withdraw the measure from consideration.
Did the Standing Committee overturn any of the Scientific Methods Panel's ratings of Scientific Acceptability? If so, state the measure and why the measure was overturned.	No	*
If a recommended measure is a related and/or competing measure, was a rationale provided for the Standing Committee's recommendation? If not, briefly explain.	No	There were no related or competing measures for the measures under review this cycle.
Were any measurement gap areas addressed? If so, identify the areas.	No	*
Are there additional concerns that require CSAC discussion? If so, briefly explain.	No	*

* Cell left intentionally blank

Appendix B: Measures Not Recommended for Endorsement

One measure was recommended for endorsement, and one was withdrawn from consideration following the post-comment meeting. No measures were not recommended.

Appendix C: NQF Member Expression of Support Results

One NQF member provided their expression of non-support for measure #3614 before it was withdrawn from consideration. Results are provided below.

NQF #3614 Hospitalization After Release With Missed Dizzy Stroke (Johns Hopkins Armstrong Institute of Patient Safety and Quality)

Member Council	Support	Do Not Support	Total
Health Professional	0	1	1

Appendix D: Details of Measure Evaluation

Rating Scale: H=High; **M**=Moderate; **L**=Low; **I**=Insufficient; **NA**=Not Applicable; **Y**=Yes; **N**=No

Vote totals may differ between measure criteria and between measures as Standing Committee members often have to join calls late or leave calls early. NQF ensures that quorum is maintained for all live voting. All voting outcomes are calculated using the number of Standing Committee members present during the meeting for that vote as the denominator. Denominator vote counts may vary throughout the criteria due to intermittent Standing Committee attendance fluctuation. The vote totals reflect members present and eligible to vote at the time of the vote. Quorum (a minimum of 12 out of 18 active non-recused Standing Committee members present) was reached and maintained for the duration of discussion and voting on #0507.

Measure Recommended for Endorsement

NQF #0507 Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports

Measure Worksheet

Description: This measure assesses the percentage of final reports for carotid imaging studies (i.e., neck magnetic resonance angiography [MRA], neck computerized tomographic angiography [CTA], neck duplex ultrasound, and carotid angiogram) performed that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement.
Numerator Statement: Final reports for carotid imaging studies that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement
Denominator Statement: All final reports for carotid imaging studies (i.e., neck MRA, neck CTA, neck duplex ultrasound, and carotid angiogram) performed
Exclusions: No denominator exclusions or denominator exceptions
Adjustment/Stratification: No risk adjustment or stratification
Level of Analysis: Clinician: Individual
Setting of Care: Outpatient Services, Inpatient/Hospital
Type of Measure: Process
Data Source: Claims, Registry Data
Measure Steward: American College of Radiology (ACR)

STANDING COMMITTEE MEETING [July 15, 2021, and July 19, 2021] & POST COMMENT MEETING [October 27, 2021]

1. Importance to Measure and Report: ***The measure meets the Importance criteria.***

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Total Votes: 15; H-0**; **M-1**; **L-0**; **I-14**

Evidence Exception: **Total Votes: 15; Yes-11; No-4 (11/15 – 73%, Pass Insufficient Evidence with Exception)**

1b. Performance Gap: **Total Votes: 14; H-0**; **M-9**; **L-3**; **I-2**

Rationale:

- Moderate and severe stenosis (50-90% occlusion) of the carotid artery affects approximately 10% of the general population by their 8th decade and causes approximately 10% of strokes.
- The stroke risk associated with asymptomatic carotid stenosis (ACS) falls between 60-80% with medical treatment alone versus additional carotid endarterectomy (CEA). This improved stroke prevention efficacy also has implications for better outcomes for patients with symptomatic carotid stenosis (SCS).
- There is a significantly higher overall risk of stroke or death associated with carotid angioplasty/stenting than with CEA.
- The Standing Committee did not see a relationship between outcomes/quality of care with accurate versus inaccurate carotid measurement. However, multiple Standing Committee members expressed that the measurement of carotid stenosis was good for improving stroke care.
- The Standing Committee ultimately voted for insufficient evidence with exception due to the

absence of empirical evidence but agreed that holding providers accountable for this measure is beneficial to patients.
- The developer provided data with performance rates increasing from 16.85 in 2012 to 74.97 in 2018.
- The Standing Committee questioned the developer's data because the original submission included data from both individual providers and groups. As the measure's level of analysis is for individuals, the group data were removed, and updated data were provided. When the group data were removed, the number of physicians included in the data analysis decreased by approximately 50%. The Standing Committee also highlighted that the number of physicians included in the data analysis decreased from about 3 million physicians in 2012 to about 9,000 in 2015 and questioned how performance could be determined.

2. Scientific Acceptability of Measure Properties: The measure meets the Scientific Acceptability criteria.

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Total Votes: 14; H-6**; **M-8**; **L-0**; **I-0**

2b. Validity: Measure Evaluation Meeting – **Total Votes: 15; H-0; M-7**; **L-5**; **I-3 (7/15, Consensus Not Reached)**; Post-Comment Meeting – **Total Votes: 12; H-0; M-9; L-2; I-1 (Pass)**

Rationale

- The beta-binominal model was used to assess the signal-to-noise ratio at the performance-score level. The overall mean reliability score was 0.9340, CI (0.99331, 0.99350).
- The Standing Committee had no concerns about reliability.
- The developer attempted construct validity by correlating the results of NQF #0507 with other measures but was unable to find suitable measures for this purpose within same accountability program.
- The developer also attempted criterion validity testing using data at the population level but was unable to format measures' data sets to perform an empirical analysis; while MIPS #409 and MIPS #413 were specified at the individual-clinician level, the Centers for Medicare & Medicaid Services (CMS) was unable to provide the developer with individual level data because all submissions were done at the group level.
- The developer performed a new face validity study in November 2020, which demonstrated that 82.15% (23 members) of the panel either agreed or strongly agreed that this measure accurately distinguishes good from poor quality.
- Empirical validity testing is required for maintenance measures. However, the Standing Committee can assess whether the developer had a reasonable approach to attempting empirical validity and whether it was sufficient to resort to face validity.
- A Standing Committee member asked the developer if they could have approached validity testing by auditing a sample number of charts outside of the measure data set to demonstrate the accuracy of the method used to report. The developer stated they did not know that this was an option for testing but that it could be possible.
- The Standing Committee did not reach consensus on validity during the measure evaluation meeting.
- During the post-comment period, the developer conducted additional data element validity testing and submitted this testing as a public comment. The Standing Committee reviewed this information and determined the measure did meet the validity criterion and passed the measure on validity.

3. **Feasibility: Total Votes: 13; H-5; M-8; L-0; I-0**

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale

- Data elements are abstracted from a record, some are in defined fields in electronic clinical data, and some elements are manually abstracted from the radiology report. The American College for Radiology (ACR) is working to enable artificial intelligence (AI) and natural language processing (NLP) in their data collection.
- Subscription fees are collected for use of this measure.

- The Standing Committee highlighted that chart abstraction is prone to errors; nonetheless, this measure has been used for a long time.
- The Standing Committee did not have any additional concerns regarding feasibility.

4. **Usability and Use: *The maintenance measure meets the Use subcriterion.***

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use: Total Votes: 13; Pass-13; No Pass-0;

4b. Usability: Total Votes: 13; H-2; M-11; L-0; I-0

Rationale

- This measure is being used for accountability (i.e., CMS Payment Program for accountability and reimbursement) and public reporting (i.e., quality improvement within ACR registries).
- This measure has created more standardization for carotid imaging results while supporting increased communications between radiologists and referring physicians.
- The Standing Committee did not identify potential harms.
- The Standing Committee did not have any concerns regarding use or usability.

5. Related and Competing Measures

- No related or competing measures were noted.

6. Standing Committee Recommendation for Endorsement: Yes-12; No-0

7. Public and Member Comment

- One comment was received from the measure developer providing additional validity testing for the Standing Committee to consider during the post-comment meeting.

Neurology
Spring 2021 Review Cycle

CSAC Review

November 30 – December 1, 2021
Funded by the Centers for Medicare & Medicaid Services under contract HHSM-500-2017-00060I Task Order HHSM-500-T0001

Neurology Standing Committee Recommendations

- **Two measures reviewed for Spring 2021**
  - One measure reviewed by the Scientific Methods Panel
    - » #3614 passed SMP on reliability and validity.
- **One measure recommended for endorsement**
  - **#0507** Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports – (American College of Radiology) maintenance
- **One measure withdrawn (following post-comment meeting)**
  - **#3614** Hospitalization After Release With Missed Dizzy Stroke (Johns Hopkins Armstrong Institute of Patient Safety and Quality) new

Overarching Issues for Neurology Measures

- **Concerns Over the Quality of the Evidence**
  - The Standing Committee questions whether better measurement of carotid stenosis itself was linked to improved outcomes. The developer was unable to describe a clear linkage between the process measure and the outcome. Instead, a two-step evidence pathway was proposed, in which improved measurement would lead to a better selection of patients for interventions, which have been shown to improve outcomes. This led the Standing Committee to vote for insufficient evidence with exception due to the lack of this linkage between process and outcome.
- **Issues With Scientific Acceptability**
  - The Standing Committee emphasized the importance of having sufficient observations to ensure the measure was reliable. Smaller hospitals may not achieve sufficient cases to generate a reliable measure.
  - For NQF #0507, concerns were raised that empirical validity testing was not conducted (i.e., measure results are compared against another valid measure of a similar concept). While the developer did attempt to conduct empirical validity testing, they were ultimately unable to because they could not find a suitable comparator measure at the same level of analysis.

Neurology: Public and Member Comment and Member Expressions of Support

- One public and member comment received
  - One from the developer providing comments related to validity testing for measure #0507

- No member expressed support or non-support

Neurology Contact Information

- NQF Project Team:
  - Tamara H. Funk, MPH, Director
  - Erin Buchanan, MPH, Manager
  - Yemsrach Kidane, PMP, Project Manager
  - Hannah Ingber, MPH, Senior Analyst
  - Sean Sullivan, MA, Coordinator
- Project Webpage: https://www.qualityforum.org/Neurology_.aspx
- Project email address: neurology@qualityforum.org

NATIONAL QUALITY FORUM
Driving measurable health improvements together

Neurology, Spring 2021 Cycle: CDP Report

DRAFT REPORT FOR CSAC REVIEW
NOVEMBER 30, 2021

This report is funded by the Centers for Medicare & Medicaid Services under contract HHSM-500-2017-00060I Task Order HHSM-500-T0001.

https://www.qualityforum.org

Contents

Executive Summary

Neurological conditions are disorders that affect the brain and the nerves. From 1990 to 2017, the Global Burden of Disease Study found the three most burdensome neurologic conditions in the United States (U.S.): stroke, Alzheimer's disease and other dementias, and migraine headache. Additionally, due to an aging population, neurological disorders are increasing in prevalence, incidence, mortality, and disability-adjusted life years (DALYs).[1] The National Quality Forum (NQF) Neurology Standing Committee oversees the measurement portfolio used to improve the quality of care for neurological conditions. This portfolio includes measures for stroke, subarachnoid and intracerebral hemorrhage, dementia, and carotid stenosis. The information regarding NQF's most recent Neurology Standing Committee meeting, as well as previous meetings, is available on NQF's project webpage.

For the spring 2021 cycle, the Standing Committee evaluated one newly submitted measure and one measure undergoing maintenance review against NQF's standard evaluation criteria.

The Standing Committee did not reach consensus on the following maintenance measure:

- **NQF #0507** Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports (American College of Radiology)

The following measure was withdrawn from consideration following the post-comment meeting:

- **NQF #3614** Hospitalization After Release With Missed Dizzy Stroke (Johns Hopkins Armstrong Institute of Patient Safety and Quality)

Overarching issues discussed by the Standing Committee included issues with evidence and measure scientific acceptability, which were concerns in both measures. Brief summaries of the measures currently under review are included in the body of the report; detailed summaries of the Standing Committee's discussion and ratings of the criteria for each measure are in Appendix A.

Introduction

Neurological conditions are disorders that affect the brain and the nerves found throughout the body and spinal cord. The Global Burden of Disease study found the three most burdensome neurological conditions in the U.S.: stroke, Alzheimer's and other dementias, and migraine headache. Additionally, the study found that due to an increasingly aging population, many neurological disorders are rising in prevalence, incidence, and mortality as well as increasing in DALYs.[1]

According to the American Academy of Neurology, a variety of quality measures exist that are related to the structure, process, and outcomes for neurological disorders, including epilepsy, child and geriatric neurology, headache, movement disorders, multiple sclerosis, neuromuscular disorders, and stroke.[2] NQF's Neurology Standing Committee assesses new and existing measures related to brain and spinal conditions brought by measure developers for endorsement.

For the spring 2021 cycle, the NQF Neurology project evaluated one new measure and one maintenance measure. The new measure, which was withdrawn from consideration following the post-comment meeting, assessed the rate of missed stroke in emergency departments (EDs) when patients present themselves with dizziness, which is a nonspecific symptom. The maintenance measure correlated to the appropriate measurement of carotid stenosis (i.e., a narrowing of the carotid artery) by radiologists on imaging studies, including computed tomography (CT), angiography, ultrasound, and magnetic resonance imaging (MRI). Proper carotid artery measurement is important in assessing eligibility for evidence-based interventions, such as carotid endarterectomy, which can reduce the risk of stroke.[3]

Carotid Stenosis

Approximately 87 percent of all strokes are ischemic strokes, in which blood flow to the brain is blocked.[4] A major cause of ischemic stroke is large blood vessel atherosclerosis, or the development of plaques in vessels, particularly the carotid arteries, which travel through the neck and brain. In 2005, the annual rate of strokes attributed to stenosis of the carotid artery was 13.4 per 100,000 persons.[5] When carotid stenosis is identified, particularly when it is causing stroke symptoms, treatment options (e.g., endarterectomy or stenting, anti-platelet medication) can be used to help reduce the risk of future stroke.[6]

NQF Portfolio of Performance Measures for Neurologic Conditions

The Neurology Standing Committee (Appendix C) oversees NQF's portfolio of Neurology measures (Appendix B), which includes measures for stroke, subarachnoid and intracerebral hemorrhage, dementia, and carotid stenosis. This portfolio contains 14 measures, and all 14 are process measures (see Table 1 below). There are no endorsed outcome, resource use, or composite measures.

Table 1. NQF Neurology Portfolio of Measures

Measure Portfolio	Process
Stroke	10*
Subarachnoid and Intracerebral Hemorrhage	2
Dementia	1

Measure Portfolio	Process
Carotid Stenosis	1
Total	14

*Six of these measures are currently NQF-endorsed with reserve status. *Reserve status* means that the measures are inactive because the last endorsement process did not identify a persistent performance gap.

Neurology Measure Evaluation

On June 21, July 15, and July 19, 2021, the Neurology Standing Committee evaluated one new measure and one measure undergoing maintenance review against NQF's standard measure evaluation criteria.

Table 2. Neurology Measure Evaluation Summary

Measure Summary	Maintenance	New	Total
Measures under consideration	1	1	2
Measures recommended for endorsement	1	0	1
Measures withdrawn	0	1	1
Reasons for not recommending	Importance – 0 Scientific Acceptability – 0 Feasibility – 0 Use and Usability – 0 Overall Suitability – 0 Competing Measure – 0	Importance – 0 Scientific Acceptability – 0 Feasibility – 0 Use and Usability – 0 Overall Suitability – 0 Competing Measure – 0	0

Comments Received Prior to Standing Committee Evaluation

NQF accepts comments on endorsed measures on an ongoing basis through the Quality Positioning System (QPS). In addition, NQF solicits comments for a continuous 16-week period during each evaluation cycle via an online tool located on the project webpage. For this evaluation cycle, the commenting period opened on April 22, 2021, and closed on June 3, 2021. Two comments were submitted and shared with the Standing Committee prior to the measure evaluation meetings (Appendix F).

Comments Received After Standing Committee Evaluation

The continuous 16-week public commenting period with NQF member support closed on September 27, 2021. Following the Committee's evaluation of the measures under review, NQF received 4 comments from 4 member organizations) and individuals pertaining to the draft report and to the measures under review (Appendix G). All comments for each measure under review have also been summarized in Appendix A.

Throughout the 16-week continuous public commenting period, NQF members had the opportunity to express their support ("support" or "do not support") for each measure submitted for endorsement consideration to inform the Committee's recommendations during the commenting period. This

expression of support (or not) during the commenting period replaces the member voting opportunity that was previously held subsequent to committee deliberations no NQF members expressed that they are in support of the measure. This information can be found in Appendix A of the post comment memo.

Overarching Issues

During the Standing Committee's discussion of the measures, two overarching issues emerged that were factored into the Standing Committee's ratings and recommendations for both measures under review and are not repeated in detail with each individual measure.

Concerns Over the Quality of the Evidence

For both measures, the Standing Committee raised concerns about the evidence presented by the developers. For outcome measures to pass the Standing Committee's review, a clear healthcare intervention that can improve the outcome must be present. For process measures to pass the Standing Committee's review, a clear linkage between the measured process and an important health outcome must also be present. The developer for the missed stroke measure (NQF #3614, an outcome measure) detailed interventions that could potentially be performed, such as physical examination maneuvers. However, the Standing Committee ultimately felt that this was not sufficient evidence to show that performing these examinations would reduce the risk of missed stroke in patients presenting with dizziness. For the carotid stenosis measure (NQF #0507, a process measure), the Standing Committee raised concerns about whether better measurement of carotid stenosis itself was linked to improved outcomes. The developer was unable to describe a clear linkage between the process measure and the outcome. Instead, a two-step evidence pathway was proposed, in which improved measurement would lead to a better selection of patients for interventions, which have been shown to improve outcomes. This led the Standing Committee to vote for insufficient evidence with exception due to the lack of this linkage between process and outcome.

Issues With Scientific Acceptability

For measures to receive NQF endorsement, they must be reliable, valid, and properly specified in order to be scientifically acceptable. Scientific acceptability concerns were raised in both measure discussions. Concerns were raised regarding the reliability of the missed stroke measure, particularly the importance of having sufficient observations to ensure the measure was reliable. Smaller hospitals may not achieve sufficient cases to generate a reliable measure. For NQF #0507, concerns were raised that empirical validity testing was not conducted (i.e., measure results are compared against another valid measure of a similar concept). While the developer did attempt to conduct empirical validity testing, they were ultimately unable to because they could not find a suitable comparator measure at the same level of analysis. The latter concern led to a "consensus not reached" vote for the measure.

Summary of Measure Evaluation

The following brief summaries of the measure evaluation highlight the major issues that the Standing Committee considered. Details of the Standing Committee's discussion and ratings of the criteria for each measure are included in Appendix A.

NQF #0507 Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports (American College of Radiology): Recommended

Description: This measure assesses the percentage of final reports for carotid imaging studies (e.g., neck magnetic resonance angiography [MRA], neck computerized tomographic angiography [CTA], neck duplex ultrasound, and carotid angiogram) performed that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement. **Measure Type**: Process; **Level of Analysis**: Clinician: Individual; **Setting of Care**: Outpatient Services, Inpatient/Hospital; **Data Source**: Claims, Registry

The Standing Committee did not vote on the recommendation for endorsement at the initial measure evaluation meeting because the Standing Committee did not reach consensus on validity—a must-pass criterion. The Standing Committee re-voted on the measure's validity at the post-comment web meeting on October 27, 2021.

The Standing Committee co-chair started the discussion by introducing the measure during the meeting on July 15, 2021 (Day 2). This process measure assesses the percentage of final reports, including neck MRA, neck CT angiogram, neck duplex ultrasound, and carotid angiogram, that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement. The developer then provided an overview of the measure. This measure is intended to standardize the way carotid stenosis is measured, considering that this measurement is important in making clinical decisions about interventions for carotid stenosis, including surgery. Decreasing the variation in measurement is important to treatment planning. This measure has been NQF-endorsed since 2013 and has been used in quality improvement and public reporting since 2011. The developer then described the details of the measure submission. It was noted that this was a maintenance measure and that empirical validity testing is required for re-endorsement. The developer explained some of the issues they faced when generating the evidence for empirical validity. After trying to compare this measure with several other measures, the developer was ultimately unsuccessful in generating empirical validity. As a replacement to the empirical validity, an additional face validity survey was performed in November 2020; in addition, 82 percent of the Expert Panel agreed or strongly agreed that the measure was beneficial for quality measurement.

The Standing Committee proceeded with discussion on the evidence criterion. NQF staff outlined their preliminary analysis and described several comments received from the Standing Committee, stating they did not observe a relationship between the measure and any objective health outcome. As a result, the measure could be considered for passing with insufficient evidence with exception. The Standing Committee co-chair clarified that the insufficient evidence with exception could be considered if the Standing Committee felt that the measure would be beneficial to hold providers accountable for this measure in the absence of empirical evidence or if better measure performance increased the accuracy of selection for carotid endarterectomy.

A Standing Committee member highlighted the importance of the level of stenosis for symptomatic carotid stenosis and that assessing the risk in asymptomatic stenosis is being tested in the Carotid Revascularization and Medical Management for Asymptomatic Carotid Stenosis Trial (CREST-2) study. Therefore, the Standing Committee member thought that the measurement of asymptomatic stenosis was less critical than the measurement of symptomatic stenosis. Another Standing Committee member

agreed it was a good idea to measure carotid stenosis for better stroke care. The co-chair added that the level of evidence supporting interventions for carotid stenosis was based on randomized trial data.

A Standing Committee member compared this measure to a previous measure evaluated by the Standing Committee about inpatient stroke mortality and questioned whether the evidence criterion assessment was different for new and maintenance measures. NQF staff clarified that the evidence criterion was assessed in the same way for both new and maintenance measures; however, for other criteria, such use being a must-pass criterion and the requirement for empiric validity were both different when evaluating maintenance measures. A separate Standing Committee member shared that they thought support for this measure would be improved with any evidence of better performance having an impact on any objective outcome, such as unnecessary surgery. The developer was then asked whether any such data existed.

The developer referred to a study in the *New England Journal of Medicine*, in which visual inspection was used to assess stenosis; this study stated that it was difficult to generate such evidence linking measurement processes to outcomes. A Standing Committee member referred to the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) criteria about levels of evidence for diagnostic studies, in which two-step inference (i.e., the pairing of diagnostic studies and treatment studies) is commonly performed. The GRADE working group has noted several pitfalls with this approach. A Standing Committee member mentioned that accepting the two-step inference for this measure would be failing to hold the measure to the same standard the Standing Committee applied to previously evaluated measures. Another Standing Committee member expressed concern that approving this measure would result in penalizing people for not including carotid measurement in imaging reports; they also expressed that this requirement felt like a “stretch” based on the data presented by the developer. The developer commented that the Measure Applications Partnership (MAP) had previously taken the following approach: If a diagnostic test was an inclusion criterion for an effective treatment approach and was assessed rigorously, the diagnostic test was validated. The Co-chair clarified that this approach had not been formalized as a rule.

The Standing Committee inquired whether the performance gap discussion would be helpful in informing the evidence discussion. NQF staff described the performance gap data, which were submitted by the developer. According to these data, the rate appeared to have increased from 17 percent from when the measure was first being used to 97.7 percent in the most recent data. The developer clarified that the data provided in the submission initially included incorrect information; as a result, updated data was provided, which showed that the performance rate for individuals is between 75 to 80 percent. A Standing Committee member questioned the value of this performance gap because the measure had not objectively shown to be useful or specifically linked to any objective outcome. In particular, there were concerns regarding whether the increase in measure performance related to improved patient care. It was also mentioned that complications related to carotid stenosis surgery had decreased over the last decade; although this improvement could not be directly linked to this measure, it was opined that it may be indirectly correlated. Ultimately, the Standing Committee voted for insufficient evidence with exception due to the absence of empirical evidence but agreed that holding providers accountable for this measure is beneficial to patients.

The Standing Committee proceeded with additional discussion regarding the performance gap criterion and requested clarification of the submitted data. The developer clarified that group data were included in the initial submission but were later removed in the updated data to align with the measure's individual level of analysis. The Standing Committee requested additional clarification about the impact of removing the group data from the data set, particularly the number of physicians who were removed from the analysis. The developer reported that by removing the group data, the number of physicians in the analysis decreased by approximately 50 percent. Additionally, a Standing Committee member highlighted that the number of physicians included in the data analysis decreased each year, starting with over 3 million in 2012 to about 9,000 in 2015; they also questioned how performance can be determined with this progressively smaller subgroup of physicians. The developer could not explain this large decline in the number of physicians because CMS provided the data set. Following this discussion, the Standing Committee voted and passed the measure on the performance gap criterion.

Next, the Standing Committee began their discussion on the scientific acceptability criteria, starting with reliability. A Standing Committee member clarified that the reliability being evaluated is the inclusion of the carotid stenosis measurement in imaging reports, not the accuracy of the measurement itself. Additionally, at the request of the Standing Committee, the developer provided additional explanation about the use of signal-to-noise ratio testing for reliability. There was no further discussion on reliability; the Standing Committee voted and passed the measure on reliability.

The Standing Committee proceeded with discussing the validity criterion and the lack of empirical validity testing, which is required for maintenance measures. The developer attempted construct validity testing by correlating results of this measure with other measures; however, they were unable to find suitable measures for this correlation within the same accountability program. The developer also attempted criterion validity testing using data at the population level but was unable to format the measures' data sets to perform empirical analysis. Additionally, the developer attempted to correlate the measure with Merit-Based Incentive Payment System (MIPS) #409 *Clinical Outcome Post Endovascular Stroke Treatment* and MIPS #413 *Door to Puncture Time for Endovascular Stroke Treatment*. However, while these measures were specified at the individual-clinician level, CMS was unable to provide the developer with the individual-level data since all submissions were done at the group level. The developer then performed a new face validity study in November 2020, which demonstrated that 82 percent of the Expert Panel either agreed or strongly agreed that this measure accurately distinguishes good from poor quality. NQF staff provided clarification about how the Standing Committee should evaluate the validity criterion, specifically that the Standing Committee should assess whether the developer had a reasonable approach to attempting empirical validity and whether it was sufficient to resort to face validity. Based upon this discussion, the Standing Committee voted and did not reach consensus on the validity criterion. A Standing Committee member asked the developer if they could have approached validity testing by auditing a sample number of charts outside of the measure's data set to demonstrate the accuracy of the method used to report. The developer stated they did not know this was an option for testing but that it could be possible. The Standing Committee strongly encouraged the inclusion of empirical validity in the next submission if endorsement is maintained.

During the meeting on July 19, 2021 (Day 3), the Standing Committee proceeded to discuss the feasibility criterion. The Standing Committee did not raise any concerns, although one comment was made about the process of chart abstraction being prone to errors. The co-chair clarified that data had been gathered on this measure for many years. It was also clarified that some data are collected from electronic fields, while others are manually abstracted from charts; in addition, it was also noted that the measure developer collects subscription fees.

The use criterion was discussed next. This measure is currently being used in a CMS accountability program and for quality improvement within the American College of Radiology (ACR) registries for public reporting. The Standing Committee raised no concerns on use.

The final Standing Committee discussion addressed the usability criteria. The Standing Committee did not identify any potential harms, and no concerns were raised.

Following the Day 3 meeting, the Standing Committee received a recording of the meeting and submitted online votes for the feasibility, use, and usability criteria. The Standing Committee passed the measure on feasibility, use, and usability. Since consensus was not reached on the validity criterion for this measure, the Standing Committee did not vote on overall suitability for endorsement.

During the public comment period, the developer submitted additional validity testing and summarized the results. Data element validity was conducted by the Committee and involved random audits of data that was submitted to the Qualified Clinicians Data Registry (QCDR) as part of the CMS MIPS program over a four-year period. Data submitted to the QCDR was compared with chart review and demonstrated a high level of concordance (98-100%) between the exam record data and registry data.

Dr. Tirschwell led the Standing Committee in a discussion of this additional validity testing. A Committee member asked how the records were chosen for the audit and the developer clarified that this was a random sample of charts that were submitted to CMS. The developer explained that what is being measured is the standard way that radiologists measure carotid stenosis, and that the overall performance rate and agreement were both high. The Standing Committee re-voted on validity and passed the measure on this criterion.

There was no additional discussion following the validity vote so the Standing Committee moved to a vote on overall suitability for endorsement. The measure passed and is recommended for endorsement.

NQF #3614 Hospitalization After Release With Missed Dizzy Stroke (Johns Hopkins Armstrong Institute for Patient Safety and Quality): Withdrawn

Description: This outcome measure tracks the rate of patients admitted to the hospital for a stroke within 30 days of being treated and released from the ED with either a nonspecific, presumed benign symptom-only dizziness diagnosis or a specific inner ear/vestibular diagnosis (collectively referred to as *benign dizziness*). The measure accounts for the epidemiologic base rate of stroke in the population under study using a risk difference approach (observed [short-term rate] minus expected [long-term rate]). **Measure Type**: Outcome; **Level of Analysis**: Facility; **Setting of Care**: Emergency Department and Services; **Data Source**: Claims

To begin the Standing Committee's discussion during the meeting on June 21, 2021 (Day 1), the Standing Committee co-chair presented an overview of the measure, which tracks the rate of patients admitted to the hospital for a stroke within 30 days of being treated and released from the ED with either a nonspecific, presumed benign symptom-only dizziness diagnosis or a specific inner ear/vestibular diagnosis (collectively referred to as benign dizziness). The outcome measure accounts for the epidemiologic base rate of stroke in the population under study using a risk difference approach (observed [short-term rate] minus expected [long-term rate].

The developer stated that in the ED, missed stroke is an important cause of serious harm to patients. The most common medical presentation for these missed strokes is patients presenting with dizziness or vertigo because it is easily mistaken for inner ear disease. Each year, approximately 45,000 to 75,000 patients with stroke present themselves to the ED with dizziness or vertigo; the diagnosis for stroke is missed, and the patients are erroneously discharged. The measure is based on the Symptom-Disease Pair Analysis of Diagnostic Error (SPADE) conceptual framework, which averages the clinically sensible and biologically possible relationship between symptom and disease to identify a likely diagnostic error. The measure is calculated using the full Medicare Fee-for-Service (FFS) data set and a three-year reporting period analogous to the time frames that are used for Medicare mortality and readmission measures. The Medicare FFS data set represents 20 percent of the sample of patients with dizziness in any given ED. The developer explained that they limited analysis to larger EDs with at least 250 dizziness treat-and-release visits per year, which relates to about 80,080 total ED visits each year. The developer also described this measure as the first operationally viable performance measure of stroke misdiagnosis for the hospital setting.

The Standing Committee proceeded with discussion on the evidence criterion. First, a Standing Committee member described the collection of data to identify dizziness and stroke diagnoses by using the International Classification of Diseases, 10th Revision (ICD-10) codes was straightforward. However, the Standing Committee member expressed concerns about the ability of an intervention being performed to improve the measure. The developer first described the intervention of the Head Impulse, Nystagmus, Test of Skew (HINTS) examination battery, which can diagnose central dangerous causes of dizziness with high sensitivity and specificity. The HINTS examination, in conjunction with positional maneuvers (e.g., the Dix-Hallpike maneuver for Benign Positional Paroxysmal Vertigo [BPPV]), can assist in the diagnosis of the causes of underlying dizziness that do not need imaging. Another Standing Committee member commented that direct evidence of the use of the HINTS examination lowering stroke rate does not appear to be presented. Rather, the literature shows that early diagnosis and interventions in stroke reduce the rates of strokes. The developer shared early clinical trial results, which showed that a Tele-Dizzy consult service was able to substantially improve the diagnosis of dizziness and eliminated the excess 30-day stroke hospitalizations from the baseline. The developer commented that consultation with a neurologist can also improve diagnostic accuracy for patients with dizziness, thus substantially increasing the yield of stroke diagnoses relative to the baseline and the accuracy of diagnoses for patients with inner ear disease. Additionally, the developer acknowledged that an area for opportunity exists among ED physicians regarding bedside diagnostic maneuvers, and the Society of Academic Emergency Medicine is currently working on guidelines for dizziness and bedside diagnostic maneuvers.

A few Standing Committee members expressed concerns with the appropriateness of the 30-day time frame and the potential for overdiagnosis. According to the developer, multiple studies show that when a patient is discharged with benign dizziness, their rate of stroke is much higher than the general population or other patients who were discharged from the ED with an unrelated diagnosis (e.g., abdominal pain). The developer also shared that they engaged a Technical Expert Panel (TEP) regarding the time frame. The developer acknowledged that a seven-day time frame could potentially provide increased precision; however, the numbers were low overall. Additionally, 30 days is often used for the Centers for Medicare & Medicaid Services' (CMS) metrics and provides face validity to ED physicians. For a 90-day time frame, it has less face validity for ED physicians, even though it is biologically accurate.

The Standing Committee also expressed concern about the incompleteness of using "dizziness" to capture all classifications of dizziness (e.g., also considering syncope, imbalance, and vertiginous). One Standing Committee member shared a recent study published in the *Academic Emergency Medicine* journal in April 2021, which showed that the HINTS examination did not identify a single central cause of stroke in 2,000 patients; there were also patients who experienced missed posterior circulation strokes. The Standing Committee member noted that this was similar to his clinical experience in which patients present themselves with intense nausea and vomiting with vertigo; however, dizziness is not a main designation, which is concerning with regard to the numerator and denominator for the measure. According to the developer, studies have shown that differentiating between different symptoms (e.g., dizziness, syncope, vertigo, unsteadiness, etc.) does not adequately differentiate between the different causes of underlying dizziness. The developer acknowledged the possibility for a patient with a missed stroke to present themselves with syncope and not be coded for dizziness. From a quality improvement standpoint, however, having a symptom-specific approach is more actionable than a more general approach to stroke misdiagnosis.

Additionally, a Standing Committee member expressed concern about the generalizability of the intervention, as few EDs have neurologists available at the bedside for consultation to perform the assessments that could have an impact on this measure. The developer highlighted that proper bedside examination is highly effective in diagnosing patients with dizziness, at accurately diagnosing peripheral inner ear diseases, and accurately diagnosing stroke. The developer also acknowledged the low accuracy of performing the examinations effectively among ED physicians; nonetheless, careful training can be effective. Additionally, the developer shared that other clinicians, in addition to neurologists (e.g., a vestibular physical therapist), may be able to assist in these examinations. The developer mentioned that there are currently no quality improvement initiatives for improving the diagnosis of dizziness because there is no way to measure performance.

The Standing Committee proceeded with a discussion on performance gap. The Standing Committee expressed recognition of the existing gap for missed stroke diagnosis. They also acknowledged the presented evidence for disparities, particularly for African Americans (i.e., greater frequency of all missed stroke as well as greater frequency of stroke) and younger patients (i.e., frequently have under-recognized stroke). The Standing Committee did not express concerns for the performance gap criterion.

The Standing Committee then began a discussion on reliability. The Scientific Methods Panel (SMP) evaluated the measure in March 2021 (SMP Meeting Summary) and passed the measure on reliability

with a moderate rating. A Standing Committee member shared the SMP's concerns about the case minimum needed for the measure and expressed concern with the measure's lack of reliability for hospitals with less than 250 cases. The Standing Committee member also questioned whether the interquartile range of 0.590 for the median reliability was sufficient. The developer reminded the Standing Committee that they used Medicare FFS data, which represent 20 percent of the overall patient population; however, if they had access to all ED discharges from every hospital, they would be able to calculate a more precise measure. The developer shared data from large and small hospitals, highlighting that hospital variability can be measured if sufficient events are available. One Standing Committee member mentioned that the use of Medicare data skews the data towards an older average age group. This led to a discussion on the appropriateness of eliminating stroke misdiagnosis among the younger population. The developer shared that the risk of stroke misdiagnosis increases sevenfold for patients between the ages of 18 to 45 compared with patients over 75 years of age. Additionally, diagnostic interventions to improve the diagnosis of dizziness has an impact on young and older patients who are inappropriately irradiated by CT when their true diagnosis is benign positional vertigo.

Due to losing sufficient attendance to continue the meeting on Day 1, the Standing Committee did not complete their discussion on the reliability criterion. After the meeting, the Standing Committee received a recording of the meeting and submitted online votes for the evidence and performance gap criteria since those discussions were completed during Day 1. The Standing Committee did not pass the measure on evidence but did pass the measure on performance gap. Evidence is a must-pass criterion; therefore, the measure was not recommended for endorsement, and no additional discussion or voting occurred.

A reconsideration request was submitted for measure #3614 by the developer contesting that the evidence submitted actually does meet the NQF Evidence criterion, and stating concerns about the fragmentation of the meeting discussion and that the lead developer was not permitted to present the measure to the Standing Committee.

During the post-comment meeting, quorum was lost before the reconsideration vote, but sufficient attendance was maintained to continue the meeting discussion. Since quorum was lost, NQF moved to open the measure for reconsideration and the Standing Committee held a full discussion of all measure criteria. For the discussion on Evidence, a Committee member commented that there needs to be an action that is tied to the outcome. The developer drew the Committee's attention to Figure 2 in their public comment which illustrated the relationship between the measure and improvement in patient outcomes, stating that the focus is on improving diagnostic accuracy in stroke, which will improve outcomes for patients in reducing morbidity and mortality.

The Chair asked the developer about whether there was an ongoing randomized trial for this measure. The developer shared that the AVERT clinical trial is ongoing and is assessing diagnostic accuracy as an outcome of a care pathway for the evaluation of patients with dizziness. Preliminary results show that experts assessing eye movements do improve diagnostic accuracy, approximately doubling the detection rate. In addition, there is evidence that the quality of treatment improves with better diagnosis. In particular, the developer stated that earlier treatment for minor stroke cuts the risk for major stroke by 34% in the next 21 days. Committee members expressed concern that this connection is

somewhat indirect, and that there is an unclear link between diagnostic accuracy and improving patient outcomes.

The Standing Committee expressed concerns with both reliability and validity. The reliability score was somewhat low. For validity, there was concern that the diagnostic codes may not accurately capture the actual miss rate. In addition, when there is a low prevalence, the positive predictive value is low, meaning subtle differences between coding and local clinical practice could confound differences in quality between hospitals observed in this measure. The Standing Committee also had questioned the usability of the measure as there was concern that neurologists may not be available to help implement this measure which could lead to unintended consequences, in particular with diametrically opposed incentives for emergency physicians to reduce diagnostic imaging and to reduce missed diagnoses.

The developer withdrew the measure from consideration after the post-comment meeting.

References

1. GBD 2017 US Neurological Disorders Collaborators, Feigin VL, Vos T, et al. Burden of Neurological Disorders Across the US From 1990-2017: A Global Burden of Disease Study. *JAMA Neurol*. 2021;78(2):165.

2. American Academy of Neurology. Quality Measures. Quality Measures. https://www.aan.com/practice/quality-measures/. Last accessed August 2021.

3. Messas E, Goudot G, Halliday A, et al. Management of carotid stenosis for primary and secondary prevention of stroke: state-of-the-art 2020: a critical review. *Eur Heart J Suppl J Eur Soc Cardiol*. 2020;22(Suppl M):M35-M42.

4. Centers for Disease Control (CDC). Types of Stroke. https://www.cdc.gov/stroke/types_of_stroke.htm. Published May 25, 2021. Last accessed August 2021.

5. Flaherty ML, Kissela B, Khoury JC, et al. Carotid Artery Stenosis as a Cause of Stroke. *Neuroepidemiology*. 2013;40(1):36-41.

6. Cleveland Clinic. Carotid Artery Disease (Carotid Artery Stenosis): Treatments. Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/16845-carotid-artery-disease-carotid-artery-stenosis. Last accessed August 2021.

Appendix A: Details of Measure Evaluation

Rating Scale: H=High; **M**=Moderate; **L**=Low; **I**=Insufficient; **NA**=Not Applicable

Vote totals may differ between measure criteria and between measures as Standing Committee members often have to join calls late or leave calls early. NQF ensures that quorum is maintained for all live voting. All voting outcomes are calculated using the number of Standing Committee members present during the meeting for that vote as the denominator. Denominator vote counts may vary throughout the criteria due to intermittent Standing Committee attendance fluctuation. The vote totals reflect members present and eligible to vote at the time of the vote. Quorum (a minimum of 12 out of 18 active non-recused Standing Committee members present) was reached and maintained for the duration of discussion and voting on #0507.

Measures Recommended

NQF #0507 Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports

Measure Worksheet | Specifications

Description: This measure assesses the percentage of final reports for carotid imaging studies (i.e., neck magnetic resonance angiography [MRA], neck computerized tomographic angiography [CTA], neck duplex ultrasound, and carotid angiogram) performed that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement.
Numerator Statement: Final reports for carotid imaging studies that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement
Denominator Statement: All final reports for carotid imaging studies (i.e., neck MRA, neck CTA, neck duplex ultrasound, and carotid angiogram) performed
Exclusions: No denominator exclusions or denominator exceptions
Adjustment/Stratification: No risk adjustment or stratification
Level of Analysis: Clinician: Individual
Setting of Care: Outpatient Services, Inpatient/Hospital
Type of Measure: Process
Data Source: Claims, Registry Data
Measure Steward: American College of Radiology (ACR)

STANDING COMMITTEE MEETING [July 15, 2021, and July 19, 2021]

1. Importance to Measure and Report: The measure meets the Importance criteria.

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Total Votes: 15; H-0**; **M-1**; **L-0**; **I-14**

Evidence Exception: **Total Votes: 15; Yes-11; No-4 (11/15 – 73%, Pass Insufficient Evidence with Exception)**

1b. Performance Gap: **Total Votes: 14; H-0**; **M-9**; **L-3**; **I-2**

Rationale

- Moderate and severe stenosis (50-90% occlusion) of the carotid artery affects approximately 10% of the general population by their 8th decade and causes approximately 10% of strokes.
- The stroke risk associated with asymptomatic carotid stenosis (ACS) falls between 60-80% with medical treatment alone versus additional carotid endarterectomy (CEA). This improved stroke prevention efficacy also has implications for better outcomes for patients with symptomatic carotid stenosis (SCS).
- There is a significantly higher overall risk of stroke or death associated with carotid angioplasty/stenting than with CEA.
- The Standing Committee did not see a relationship between outcomes/quality of care with accurate versus inaccurate carotid measurement. However, multiple Standing Committee

members expressed that the measurement of carotid stenosis was good for improving stroke care.

- The Standing Committee ultimately voted for insufficient evidence with exception due to the absence of empirical evidence but agreed that holding providers accountable for this measure is beneficial to patients.
- The developer provided data with performance rates increasing from 16.85 in 2012 to 74.97 in 2018.
- The Standing Committee questioned the developer's data because the original submission included data from both individual providers and groups. As the measure's level of analysis is for individuals, the group data were removed, and updated data were provided. When the group data were removed, the number of physicians included in the data analysis decreased by approximately 50%. The Standing Committee also highlighted that the number of physicians included in the data analysis decreased from about 3 million physicians in 2012 to about 9,000 in 2015 and questioned how performance could be determined.

2. Scientific Acceptability of Measure Properties: The measure is "consensus not reached" on the Scientific Acceptability criteria.

(2a. Reliability precise specifications, testing; 2b. Validity testing, threats to validity)

2a. Reliability: **Total Votes: 14; H-6**; **M-8**; **L-0**; **I-0**

2b. Validity: Measure Evaluation Meeting – **Total Votes: 15; H-0;M-7**; **L-5**; **I-3 (7/15 – 47%, Consensus Not Reached)**; Post-Comment Meeting – **Total Votes: 12; H-0;M-9; L-2; I-1 (9/12 – Pass)**

Rationale

- The beta-binominal model was used to assess the signal-to-noise ratio at the performance-score level. The overall mean reliability score was 0.9340, CI (0.99331, 0.99350).
- The Standing Committee had no concerns about reliability.
- The developer attempted construct validity by correlating the results of NQF #0507 with other measures but was unable to find suitable measures for this purpose within same accountability program.
- The developer also attempted criterion validity testing using data at the population level but was unable to format measures' data sets to perform an empirical analysis; while MIPS #409 and MIPS #413 were specified at the individual-clinician level, the Centers for Medicare & Medicaid Services (CMS) was unable to provide the developer with individual level data because all submissions were done at the group level.
- The developer performed a new face validity study in November 2020, which demonstrated that 82.15% (23 members) of the panel either agreed or strongly agreed that this measure accurately distinguishes good from poor quality.
- Empirical validity testing is required for maintenance measures. However, the Standing Committee can assess whether the developer had a reasonable approach to attempting empirical validity and whether it was sufficient to resort to face validity.
- A Standing Committee member asked the developer if they could have approached validity testing by auditing a sample number of charts outside of the measure data set to demonstrate the accuracy of the method used to report. The developer stated they did not know that this was an option for testing but that it could be possible.
- The Standing Committee did not reach consensus on validity during the measure evaluation meeting.
- During the post-comment period, the developer conducted additional data element validity testing and submitted this testing as a public comment. The Standing Committee reviewed this information and determined the measure did meet the validity criterion and passed the measure on validity.

3. Feasibility: Total Votes: 13; H-5; M-8; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale:

- Data elements are abstracted from a record, some are in defined fields in electronic clinical data, and some elements are manually abstracted from the radiology report. The American College for Radiology (ACR) is working to enable artificial intelligence (AI) and natural language processing (NLP) in their data collection.
- Subscription fees are collected for use of this measure.
- The Standing Committee highlighted that chart abstraction is prone to errors; nonetheless, this measure has been used for a long time.
- The Standing Committee did not have any additional concerns regarding feasibility.

4. Usability and Use: The maintenance measure meets the Use subcriterion.

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use**: Total Votes: 13; Pass-13; No Pass-0**

4b. Usability: **Total Votes: 13; H-2; M-11; L-0; I-0**

Rationale

- This measure is being used for accountability (i.e., CMS Payment Program for accountability and reimbursement) and public reporting (i.e., quality improvement within ACR registries).
- This measure has created more standardization for carotid imaging results while supporting increased communications between radiologists and referring physicians.
- The Standing Committee did not identify potential harms.
- The Standing Committee did not have any concerns regarding use or usability.

5. Related and Competing Measures

- No related or competing measures were noted.

6. Standing Committee Recommendation for Endorsement: Yes-12; No-0

7. Public and Member Comment

- One comment was received from the measure developer providing additional validity testing for the Standing Committee to consider during the post-comment meeting.

Measures Withdrawn

NQF #3614 Hospitalization After Release With Missed Dizzy Stroke

Measure Submission

Description: This outcome measure tracks the rate of patients admitted to the hospital for a stroke within 30 days of being treated and released from the ED with either a nonspecific, presumed benign symptom-only dizziness diagnosis or a specific inner ear/vestibular diagnosis (collectively referred to as *benign dizziness*). The measure accounts for the epidemiologic base rate of stroke in the population under study using a risk difference approach (observed [short-term rate] minus expected [long-term rate]).

Numerator Statement: The number of ED index visits during the performance period that are followed within 30 days by an inpatient hospital admission to any hospital that ends in a primary diagnosis of stroke

Denominator Statement: Patients discharged from the ED with benign dizziness as the primary diagnosis code, counting a patient's first such discharge during the performance period (an "index visit") and all subsequent such discharges that fall outside a 360-day follow-up window from the previous qualifying "index visit"

Exclusions: No exclusions

Adjustment/Stratification: No risk adjustment or stratification

Level of Analysis: Facility

Setting of Care: Emergency Department and Services

Type of Measure: Outcome

Data Source: Claims

Measure Steward: Johns Hopkins Armstrong Institute of Patient Safety and Quality

STANDING COMMITTEE MEETING [June 21, 2021]

1. Importance to Measure and Report: The measure does not meet the Importance criteria.

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Measure Evaluation Meeting Total Votes: 13; Yes-4; No-9**

Post-Comment Meeting Total Votes: 14; Yes-5; No-10

1b. Performance Gap: **Total Votes: 13; H-3**; **M-8**; **L-2**; **I-0**

Rationale

- This is intended to be a measure of patients who had a treat-and-release ED visit with a diagnosis of benign dizziness who were discharged and later had a stroke, with the suggestion being that the dizziness treat-and-release ED visit reflected a potentially missed stroke diagnosis.
- Dizziness is commonly misdiagnosed in the ED with rates as high as 80%.
- Patients hospitalized for stroke (about 190,000 admissions from 9 U.S. states in 2009) are more likely to have had a treat-and-release ED visit for so-called "benign" dizziness within the prior 14 days than those who have had an ED visit for a different chief complaint.
- Benign dizziness treat-and-release discharges from the ED (about 30,000 visits per year) are more likely to return for an inpatient stroke admission within the subsequent 30 days than a heart attack admission.
- The Standing Committee expressed concerns about the ability of an intervention (e.g., Head Impulse, Nystagmus, Test of Skew [HINTS] examination or positional maneuvers, such as Dix-Hallpike for Benign Positional Paroxysmal Vertigo) being performed to improve the measure as well as the lack of evidence to support that these particular interventions lower stroke rates.
- Some Standing Committee members expressed concern about the appropriateness of the 30-day time frame and a concern about the potential for overdiagnosis. The developer mentioned that they engaged a Technical Expert Panel (TEP) regarding the time frame, and overall, a period of 30 days provided face validity to ED physicians as is often used for other CMS metrics.
- The Standing Committee also expressed concern about the incompleteness of using "dizziness" to capture all classifications of dizziness (e.g., also considering syncope, imbalance, and vertiginous) and referenced an April 2021 journal article that showed that the HINTS examination did not identify a single central cause of stroke in 2,000 patients as well as patients who experienced missed posterior circulation strokes. According to the developer, studies have shown that differentiating between different symptoms (e.g., dizziness, syncope, vertigo, unsteadiness, etc.) does not adequately differentiate between the different causes of underlying dizziness.
- The Standing Committee additionally expressed concern about the generalizability of the intervention, as few EDs have neurologists available at the bedside for consultation to perform the assessments that could have an impact on this measure. The developer highlighted that proper bedside examination is highly effective in diagnosing patients with dizziness, at accurately diagnosing peripheral inner ear diseases, and accurately diagnosing stroke.
- The Standing Committee had no concerns regarding performance gap.
- Quorum was not present during the meeting, and online voting occurred for the evidence and performance gap criteria. The Standing Committee did not pass the measure on evidence; therefore, the measure was not recommended for endorsement.
- The developer submitted a reconsideration request for this measure. During the post-comment meeting, quorum was lost prior to voting on reconsideration but sufficient attendance was maintained to continue discussion. Since quorum was lost, NQF moved to open the measure for reconsideration and the Standing Committee held a full discussion of all measure criteria, including Evidence.
- The Standing Committee expressed continued concerns that the evidence supporting this outcome was indirect, and that the link between diagnostic accuracy and improving patient outcomes is still unclear.

- Following the meeting, the Standing Committee received a recording of the meeting and voted online voting tool. The Standing Committee again did not pass the measure on evidence; therefore, the measure was not recommended for endorsement. The developer withdrew the measure from consideration following the post-comment meeting.

2. Scientific Acceptability of Measure Properties: *The measure meets the Scientific Acceptability criteria*

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **Vote Not Taken**; 2b. Validity: **Vote Not Taken**

Rationale

- In their preliminary analyses, the SMP noted minor concerns with the reliability of this measure and passed it with a moderate rating (Total votes = 8; H-0; M-5; L-1; I-2)
- For validity, the SMP expressed concerns regarding the data element validity for both the numerator and denominator, the risk adjustment methodology, and whether the measure would show meaningful differences between hospitals, given the need for a three-year measurement period and how providers would operationalize it in quality improvement efforts.
- The SMP passed the measure on validity (Total votes -8; H-0; M-5; L-2; I-1).
- The Standing Committee did not vote on the remaining criteria because the measure did not pass on evidence and the measure was not recommended for endorsement.

3. Feasibility: Vote Not Taken

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified; 3d. Data collection strategy can be implemented)

Rationale: N/A

4. Usability and Use: *The maintenance measure meets the Use subcriterion*

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

4a. Use: **Vote Not Taken** 4b. Usability: **Vote Not Taken**

Rational: N/A

5. Related and Competing Measures

- No related or competing measures noted.

6. Standing Committee Recommendation for Endorsement: Total votes: XX; Yes-X; No-X

7. **Public and Member Comment**

- The measure developer submitted an additional presentation of evidence to support a re-discussion of this criterion and to help enhance the Standing Committee's understanding of the evidence in the original submission.
- Two comments were supportive of this measure and asked the Standing Committee to reconsider its decision on Evidence.
- Two comments were not in support. One expressed concerns with the Evidence and Scientific Acceptability of the measure, and the other expressed concerns about the lack of exclusions, delineation of a case minimum, and risk adjustment, as well as the lack of variation in performance scores.
- During the post-comment discussion, the Standing Committee expressed additional concerns with both reliability and validity. The reliability score was somewhat low. For validity, there was concern that the diagnostic codes may not accurately capture the actual miss rate. In addition, when there is a low prevalence, the positive predictive value is low, meaning subtle differences between coding and local clinical practice could confound differences in quality between hospitals observed in this measure. The Standing Committee also had questioned the usability of the measure as there was concern that neurologists may not be available to help implement this measure which could lead to unintended consequences, in particular with diametrically opposed incentives for emergency physicians to reduce diagnostic imaging and to reduce missed diagnoses.

Appendix B: Neurology Portfolio—Use in Federal Programs[a]

NQF #	Title	Federal Programs: Finalized or Implemented as of June 30, 2021
0434*	STK 01: Venous Thromboembolism (VTE) Prophylaxis	N/A
0435*	STK 02: Discharged on Antithrombotic Therapy	Hospital Inpatient Quality Reporting (Implemented 2015) Medicare and Medicaid Promoting Interoperability Program for Eligible Hospitals and Critical Access Hospitals (Implemented 2012)
0436*	STK 03: Anticoagulation Therapy for Atrial Fibrillation/Flutter	Hospital Inpatient Quality Reporting (Implemented 2015) Medicare and Medicaid Promoting Interoperability Program for Eligible Hospitals and Critical Access Hospitals (Implemented 2012)
0437	STK 04: Thrombolytic Therapy	N/A
0438*	STK 05: Antithrombotic Therapy by End of Hospital Day Two	Hospital Inpatient Quality Reporting (Implemented 2015) Medicare and Medicaid Promoting Interoperability Program for Eligible Hospitals and Critical Access Hospitals (Implemented 2012)
0439*	STK 06: Discharged on Statin Medication	Hospital Inpatient Quality Reporting (Implemented 2015) Medicare and Medicaid Promoting Interoperability Program for Eligible Hospitals and Critical Access Hospitals (Implemented 2012)
0441*	STK 10: Assessed for Rehabilitation	N/A
0507	Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports	Merit-Based Incentive Payment System (MIPS) Program (Implemented 2018) Physician Compare (Implemented 2018)
0661	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	Care Compare (Implemented 2016) Hospital Outpatient Quality Reporting (Implemented 2012)
1952	Time to Intravenous Thrombolytic Therapy	N/A

[a] Per CMS Measures Inventory Tool as of June 30, 2021

NQF #	Title	Federal Programs: Finalized or Implemented as of June 30, 2021
2863	CSTK-06: Nimodipine Treatment Administered	N/A
2864	CSTK-01: National Institutes of Health Stroke Scale (NIHSS) Score Performed for Ischemic Stroke Patients	Merit-Based Incentive Payment System (MIPS) Program (Implemented 2018)
2866	CSTK-03: Severity Measurement Performed for Subarachnoid Hemorrhage (SAH) and Intracerebral Hemorrhage (ICH) Patients (Overall Rate)	N/A
2872e	Dementia: Cognitive Assessment	Medicaid Promoting Interoperability Program for Eligible Professionals (Implemented 2019) Merit-Based Incentive Payment System (MIPS) Program (Implemented 2018) Physician Compare (Implemented 2018)

*Endorsed With Reserve Status

Appendix C: Neurology Standing Committee and NQF Staff

STANDING COMMITTEE

David Tirschwell, MD, MSc (Co-Chair)
University of Washington, Harborview Medical Center
Seattle, Washington

Mary Kay Ballasiotes (inactive)
Executive Director, International Alliance for Pediatric Stroke
Charlotte, North Carolina

Jocelyn Bautista, MD
Cleveland Clinic Neurological Institute Epilepsy Center
Cleveland, Ohio

James Burke, MD
University of Michigan
Ann Arbor, Michigan

Valerie Cotter, DrNP, AGPCNP-BC, FAANP
John Hopkins School of Nursing
Baltimore, Maryland

Rebecca Desrocher, MS
Deputy Director, Health Resources and Services Administration (HRSA)
Rockville, Maryland

Bradford Dickerson, MD, MMSC
Massachusetts General Hospital
Charleston, Massachusetts

Dorothy Edwards, PhD
Director, Collaborative Center for Health Equity, University of Wisconsin Madison School of Medicine and Public Health
Madison, Wisconsin

Reuven Ferziger, MD
Director, U.S. Medical Affairs, Merck and Company
Silver Spring, Maryland

Susan Fowler, RN, PhD, CNRN, FAHA
Associate Professor, Chamberlain College of Nursing – New Jersey
Metuchen, New Jersey

Edward Jauch, MD, MS
Chief of System Research, Mission Research Institute
Asheville, North Carolina

Charlotte Jones, MD, PhD, MSPH

U.S. Food and Drug Administration (FDA)
Silver Spring, Maryland

Scott Mendelson, MD, PhD
Assistant Professor and Chief Quality Officer, University of Chicago, Department of Neurology
Chicago, Illinois

David Newman-Toker, MD, PhD
Professor of Neurology and Director, AI Center for Diagnostic Excellence, Armstrong Institute for Patient Safety and Quality at Johns Hopkins University
Baltimore, Maryland

Melody Ryan, PharmD, MPH
University of Kentucky College of Pharmacy
Lexington, Kentucky

Michael Schneck, MD
Professor of Neurology and Neurosurgery, Loyola University Medical Center
Maywood, Illinois

Jane Sullivan, PT, DHS, MS
Northwestern University
Chicago, Illinois

Kelly Sullivan, PhD
Georgia Southern University
Statesboro, Georgia

Max Wintermark, MD, MS
Professor of Radiology and Chief of Neuroradiology, Stanford University
Stanford, California

Ross Zafonte, DO
Harvard Medical School
Boston, Massachusetts

NQF STAFF

Kathleen F. Giblin
Acting Senior Vice President, Quality Measurement

Tricia Elliott, MBA, CPHQ, FNAHQ
Senior Managing Director, Quality Measurement

Tamara H. Funk, MPH
Director

Erin Buchanan, MPH

Manager

Hannah Ingber, MPH
Senior Analyst

Sean Sullivan, MA
Coordinator

Yemsrach Kidane, PMP
Project Manager

Chelsea Lynch, MPH, MSN, RN, CIC
Director

Oroma Igwe, MPH
Manager

Monika Harvey, MBA, PMP
Project Manager

Jonah Lewis
Administrative Assistant

Jesse Pines, MD, MBA, MSCE
Consultant

Appendix D: Measure Specifications

#0507 Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports

STEWARD

American College of Radiology (ACR)

DESCRIPTION

Percentage of final reports for carotid imaging studies (neck magnetic resonance angiography (MRA), neck computerized tomographic angiography (CTA), neck duplex ultrasound, carotid angiogram) performed that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement

TYPE

Process

DATA SOURCE

Claims, Registry Data Not applicable

LEVEL

Clinician: Individual

SETTING

Inpatient/Hospital, Outpatient Services

NUMERATOR STATEMENT

Final reports for carotid imaging studies that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement

NUMERATOR DETAILS

Definition:

"Direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement" includes direct angiographic stenosis calculation based on the distal lumen as the denominator for stenosis measurement OR an equivalent validated method referenced to the above method (e.g., for duplex ultrasound studies, velocity parameters that correlate with anatomic measurements that use the distal internal carotid lumen as the denominator for stenosis measurement).

Numerator Instructions:

For duplex imaging studies the reference is indirect, since the degree of stenosis is inferred from velocity parameters and cross referenced to published or self-generated correlations among velocity parameters and results of angiography or other imaging studies which serve as the gold standard. In Doppler ultrasound, the degree of stenosis can be estimated using Doppler parameter of the peak systolic velocity (PSV) of the internal

carotid artery (ICA), with concordance of the degree of narrowing of the ICA lumen. Additional Doppler parameters of ICA-to-common carotid artery (CCA) PSV ratio and ICA end-diastolic velocity (EDV) can be used when degree of stenosis is uncertain from ICA PSV. (Grant et al, 2003)

Measure performance is met when study methodology is identified and findings are reported as a percentage or range of percentages of carotid stenosis. Documented findings of “No Stenosis” determined through NASCET or comparable methodology also meet measure performance. A short note can be made in the final report, such as:

A short note can be made in the final report, such as:

- "Severe left ICA stenosis of 70-80% by NASCET criteria” or
- “Severe left ICA stenosis of 70-80% by criteria similar to NASCET” or
- “70% stenosis derived by comparing the narrowest segment with the distal luminal diameter as related to the submitted measure of arterial narrowing” or
- “Severe stenosis of 70-80% - validated velocity measurements with angiographic measurements, velocity criteria are extrapolated from diameter data as defined by the Society of Radiologists in Ultrasound Consensus Conference Radiology 2003; 229;340-346”.

In a small number of denominator cases the distal ICA may not be viewed (e.g. an innominate artery or common carotid injection). Performance would be met if there is documentation, for example, that indicates “stenosis measurements are made with reference to the distal lumen”, as a matter of process and consistent practice method.

DENOMINATOR STATEMENT

All final reports for carotid imaging studies (neck MRA, neck CTA, neck duplex ultrasound, carotid angiogram) performed

DENOMINATOR DETAILS

This measure is to be submitted each time a carotid imaging study is performed during the performance period for all patients, regardless of age. There is no diagnosis associated with this measure. Eligible clinicians who provide the professional component of diagnostic imaging studies of the carotids will submit this measure.

Denominator Criteria (Eligible Cases) for Claims and Registry:

Patient procedure during the performance period (CPT): 36221, 36222, 36223, 36224, 37215, 37216*, 37217, 37218, 70498, 70547, 70548, 70549, 93880, 93882

DENOMINATOR NOTE: (*) Signifies that this CPT Category I code is a non-covered service under the Medicare Part B Physician Fee Schedule (PFS). These non-covered services should be counted in the denominator population for MIPS CQMs

EXCLUSIONS

No Denominator Exclusions or Denominator Exceptions

EXCLUSION DETAILS

None

RISK ADJUSTMENT

No risk adjustment or risk stratification

STRATIFICATION

We encourage the results of this measure to be stratified by race, ethnicity, gender, and primary language, and have included these variables as recommended data elements to be collected.

TYPE SCORE

Rate/proportion/better quality = higher score

ALGORITHM

To calculate performance rates:

1) Find the patients who meet the initial patient population (i.e., the general group of patients that the performance measure is designed to address).
2) From the patients within the initial patient population criteria, find the patients who qualify for the denominator (i.e., the specific group of patients for inclusion in a specific performance measure based on defined criteria). Note: in some cases the initial patient population and denominator are identical.
3) From the patients within the denominator, find the patients who qualify for the Numerator (i.e., the group of patients in the denominator for whom a process or outcome of care occurs). Validate that the number of patients in the numerator is less than or equal to the number of patients in the denominator.

If the patient does not meet the numerator, this case represents a quality failure. 108475| 145989| 141015| 142351| 151468

COPYRIGHT / DISCLAIMER

Appendix E: Related and Competing Measures

There are no related or competing measures for the measures evaluated by the Neurology Standing Committee during the spring 2021 cycle.

Appendix F: Pre-Evaluation Comments

Comments received as of July 16, 2021.

#3614 Hospitalization After Release With Missed Dizzy Stroke (H.A.R.M Dizzy-Stroke)

COMMENTER

Submitted by Donald May, Federation of American Hospitals

COMMENT

The Federation of American Hospitals (FAH) appreciates the opportunity to comment on this measure. While FAH supports the measure's focus of driving improvements in diagnostic accuracy, we are concerned that the measure may require additional exclusions and question if case minimums to ensure adequate reliability and risk adjustment are needed and whether the measure scores produce sufficient variation to make the results meaningful for accountability purposes. The FAH asks the Standing Committee to consider whether some exclusions, delineation of a case minimum, and possible risk adjustment would be appropriate for inclusion in this measure. For example, is it appropriate to hold a facility accountable for a possible missed diagnosis when an individual leaves against medical advice (AMA)? We are also concerned that a minimum number of patients will be required to ensure that the measure produces acceptable reliability thresholds of 0.7 or higher, yet we were unable to identify any such requirement. Finally, while we appreciate the analyses completed to justify the lack of risk adjustment, we request that the Committee discuss whether there are any clinical or social risk factors that could contribute to an individual presenting with a stroke within the 30-day window that is unrelated to the chief complaint of dizziness during the emergency department visit and as a result if there should be some adjustment based on those variables. The FAH also questions the usefulness of this measure given the limited variation in performance scores with no hospitals identified as statistically worse than the national average; only eight were identified as having significant harm, and the vast majority of the hospitals were no different or better than the national average. We do not believe that this measure provides any new information that would be useful to hospitals and patients. The FAH asks that the Committee carefully consider these concerns during their review.

#3614 Hospitalization After Release With Missed Dizzy Stroke (H.A.R.M Dizzy-Stroke)

COMMENTER

Submitted by Koryn Rubin

COMMENT

The American Medical Association (AMA) appreciates the opportunity to comment on this measure. While addressing diagnostic error is absolutely critical to ensuring that patients receive the highest quality of care possible, we are concerned with the lack of exclusions such as those patients who leave against medical advice and question whether the measure should be risk adjusted for clinical and/or social risk factors. Specifically, it remains unclear to us whether there are any factors that could contribute to an individual being treated for benign dizziness

but then present with an unrelated stroke within the 30-day time window and if this scenario is possible, we believe that the measure should include risk adjustment.

In addition, we are disappointed to see the minimum measure score reliability results appeared to be less than 0.2 according to the histogram included in the testing form. While the median reliability score was 0.590, we believe that measures must meet minimum acceptable thresholds of 0.7 for reliability and the developer should include a minimum case number as a part of the measure specifications to achieve this threshold across all reporting hospitals.

Lastly, we question whether the information provided as a result of this measure is truly useful for accountability purposes and for informing patients on the quality of care provided by hospitals. Specifically, our concern relates to the relatively limited amount of variation across facilities. While 627 hospitals out of the 967 facilities were identified as performing "Better" than the national average, zero hospitals performed "Worse" and only eight were identified as having statistically significant "Harm". Endorsing a measure that currently only identifies such a small number of outliers does not enable users to distinguish meaningful differences in performance and limits a measure's usability.

We request that the Standing Committee evaluate whether the measure adequately meets the measure evaluation criteria.

Appendix G: Post-Evaluation Comments

NQF #0507 Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports, Comment #7799

Standing Committee Recommendation: Consensus Not Reached

Comment ID#: 7799

Commenter: Submitted by Karen Orozco

Council / Public: HPR

Comment Period: Post-Evaluation Public and Member Commenting

Date Comment was Submitted: 9/23/2021

Developer Response Required? No

Level of Support: Member Does not support

Theme:

Comment

The American College of Radiology (ACR) has completed data element empirical validity to support NQF 0507 re-endorsement.

According to the Blueprint for the Centers for Medicare & Medicaid Services (CMS) Measures Management System, data element validity is the "extent to which the information represented by the data element or code used in the measure reflects the actual concept or event intended." The ACR performed random audits using the groups that submitted Qualified Clinical Data Registry (QCDR) records for the measure to CMS for their Merit-based Incentive Payment System (MIPS) program, over a four-year period. The auditors compared the numerator data element registry submissions used in measure calculation with actual exam records from the submitters' systems. The audit confirmed a high level of agreement and concordance between the data shown on exam records and what was submitted to the registry. The records where exam data did not match the registry data represent human error in collection or submission of data.

Summary of NQF 0507 in 2017-2020 Audit by Year

Year	# Groups	# Records Audited	# Records Without Issue	% Records Without Issue
2017	11	108	106	98%
2018	22	128	126	98%
2019	17	130	130	100%
2020	15	69	69	100%

Developer Response

N/A

NQF Response

N/A

NQF Committee Response

Thank you for your comment. The Standing Committee reviewed and considered this information during the post-comment meeting and found the measure's validity acceptable and the measure to be suitable for endorsement at this time.

NQF #3614 Hospitalization After Release with Missed Dizzy Stroke (H.A.R.M Dizzy-Stroke), Comment #7756

Standing Committee Recommendation: Measure Not Recommended for Endorsement

Comment ID#: 7756

Commenter: Submitted by David Morrill

Council / Public: Public

Comment Period: Post-Evaluation Public and Member Commenting

Date Comment was Submitted: 9/7/2021

Developer Response Required? No

Level of Support: N/A

Theme: N/A

Comment

As a patient who struggles with ongoing vestibular issues as a result of a missed stroke diagnosis, I am deeply invested in supporting efforts to improve the diagnostic performance of Emergency Department (ED) physicians in identifying strokes in patients who present with symptoms of dizziness and vertigo. I am writing to urge the NQF Neurology Standing Committee to reconsider its vote on the "Evidence" criteria for measure #3614"Hospitalization After Release with Missed Dizzy Stroke (H.A.R.M Dizzy-Stroke)".

Based on my work with the Vestibular Disorder Association (VeDA), I have been made aware of a number of diagnostic techniques that ED physicians can use to improve their performance in diagnosing patients that present with dizziness. These techniques include the Head Impulse, Nystagmus and Test of Skew (HINTS) exam; ordering MRIs instead of CT scans; and following published specialty guidelines (e.g SAEMS' GRACE3 guideline). As I understand the current healthcare landscape, these techniques have not been put into wide use.

Without a measure of diagnostic performance, the underlying assumption is that current diagnostic approaches are yielding an adequate result (and they're not!). Measure #3614 would provide the "feedback" loop this is currently missing on how well ED physicians recognize dizziness as "benign", as opposed to a symptom of a potentially devastating outcome, such as stroke.

My own misdiagnosis has left me with a life-long disability. Given the availability of interventions to improve diagnosis of dizzy patients, I would strongly urge the Standing Committee to reconsider its earlier vote on the "Evidence" associated with this measure and to support the measure's full endorsement.

David M Morrill Stroke Patient

Developer Response

N/A

NQF Response

Thank you for your comment.

NQF Committee Response

Thank you for your comment. The Standing Committee reviewed and discussed this information during the post-comment meeting but determined that the Evidence was not sufficient at this time to pass the measure.

NQF #3614 Hospitalization After Release with Missed Dizzy Stroke (H.A.R.M Dizzy-Stroke), Comment #7806

Standing Committee Recommendation: Measure Not Recommended for Endorsement

Comment ID#: 7806

Commenter: Submitted by Koryn Rubin

Council / Public: Health Professional

Comment Period: Post-Evaluation Public and Member Commenting

Date Comment was Submitted: 9/24/2021

Developer Response Required? No

Level of Support: Member Does not support

Theme: N/A

Comment

The American Medical Association (AMA) agrees with the concerns raised by the Standing Committee on this measure, particularly around the scientific acceptability of the measure. We support the Committee's recommendation to not endorse the measure at this time.

Developer Response

N/A

NQF Response

Thank you for your comment.

NQF Committee Response

Thank you for your comment. The Standing Committee reviewed and discussed this information and the developer's response during the post-comment meeting and agreed with many of the concerns around scientific acceptability. Ultimately, the Standing Committee determined that the Evidence was not sufficient at this time to pass the measure.

NQF #3614 Hospitalization After Release with Missed Dizzy Stroke (H.A.R.M Dizzy-Stroke), Comment #7752

Standing Committee Recommendation: Measure Not Recommended for Endorsement

Comment ID#: 7752

Commenter: Submitted by David Morrill

Council / Public: Public

Comment Period: Post-Evaluation Public and Member Commenting

Date Comment was Submitted: 9/7/2021

Developer Response Required? No

Level of Support: N/A

Theme: Supportive of measure

Comment

I am deeply invested in supporting efforts to improve the diagnostic performance of Emergency Department (ED) physicians in identifying strokes in patients who present with symptoms of dizziness and vertigo. I am writing to urge the NQF Neurology Standing Committee to reconsider its vote on the "Evidence" criteria for measure #3614"Hospitalization After Release with Missed Dizzy Stroke (H.A.R.M Dizzy-Stroke)".

Based on my work with the Vestibular Disorder Association (VeDA), I have been made aware of a number of diagnostic techniques that ED physicians can use to improve their performance in diagnosing patients that present with dizziness. These techniques include the Head Impulse, Nystagmus and Test of Skew (HINTS) exam; ordering MRIs instead of CT scans; and following published specialty guidelines (e.g SAEMS' GRACE3 guideline). As I understand the current healthcare landscape, these techniques have not been put into wide use.

Without a measure of diagnostic performance, the underlying assumption is that current diagnostic approaches are yielding an adequate result (and they're not!). Measure #3614 would provide the "feedback" loop this is currently missing on how well ED physicians recognize dizziness as "benign", as opposed to a symptom of a potentially devastating outcome, such as stroke.

My own misdiagnosis has left me with a life-long disability. Given the availability of interventions to improve diagnosis of dizzy patients, I would strongly urge the Standing Committee to reconsider its earlier vote on the "Evidence" associated with this measure and to support the measure's full endorsement.

Developer Response

N/A

NQF Response

Thank you for your comment.

NQF Committee Response

N/A

NQF #3614 Hospitalization After Release with Missed Dizzy Stroke (H.A.R.M Dizzy-Stroke), Comment #7641

Standing Committee Recommendation: Measure Not Recommended for Endorsement

Comment ID#: 7641

Commenter: Submitted by Donald May, Federation of American Hospitals

Council / Public: PRO

Comment Period: Post-Evaluation Public and Member Commenting

Date Comment was Submitted: 6/10/2021

Developer Response Required? Yes

Level of Support: N/A

Theme: N/A

Comment

#3614, Hospitalization After Release with Missed Dizzy Stroke (H.A.R.M Dizzy-Stroke)The Federation of American Hospitals (FAH) appreciates the opportunity to comment on this measure. While FAH supports the measure's focus of driving improvements in diagnostic accuracy, we are concerned that the measure may require additional exclusions and question if case minimums to ensure adequate reliability and risk adjustment are needed and whether the measure scores produce sufficient variation to make the results meaningful for accountability purposes. The FAH asks the Standing Committee to consider whether some exclusions, delineation of a case minimum, and possible risk adjustment would be appropriate for inclusion in this measure. For example, is it appropriate to hold a facility accountable for a possible missed diagnosis when an individual leaves against medical advice (AMA)? We are also concerned that a minimum number of patients will be required to ensure that the measure produces acceptable reliability thresholds of 0.7 or higher, yet we were unable to identify any such requirement. Finally, while we appreciate the analyses completed to justify the lack of risk adjustment, we request that the committee discuss whether there are any clinical or social risk

factors that could contribute to an individual presenting with a stroke within the 30-day window that is unrelated to the chief complaint of dizziness during the emergency department visit and as a result if there should be some adjustment based on those variables. The FAH also questions the usefulness of this measure given the limited variation in performance scores with no hospitals identified as statistically worse than the national average, only 8 were identified as having significant harm and the vast majority of the hospitals were no different or better than the national average. We do not believe that this measure provides any new information that would be useful to hospitals and patients. The FAH asks that the committee carefully consider these concerns during their review.

Developer Response

We appreciate the opportunity to respond to The Federation of American Hospitals' (FAH) comments on measure #3614 under review by the NQF Neurology Standing Committee.

The concerns raised by FAH primarily relate to the scientific acceptability of the measure. These aspects of the measure have already been reviewed and discussed by the NQF Scientific Methods Panel, where the panel voted to pass the measure on Scientific Acceptability. We will address FAH's comments in brief below and would urge Standing Committee members and other interested parties to review the Scientific Methods Panel meeting notes for additional detail about these topics.

Lack of exclusions: Patients who left against medical advice (AMA) were excluded. We apologize for any lack of clarity on this point in the documentation. We are happy to provide additional information on this issue if the Committee so desires.

Minimum sample size for reliability: As described in our submitted testing documentation, we restricted our sample to those hospital EDs that had at least 250 "benign dizziness" discharges from the ED during the 3-year performance period (i.e., the measure denominator needs to be 250 or higher). The median reliability score for the entire 967 hospital sample was 0.590, with an interquartile range of 0.414-0.951. These values closely mirror the reliability statistics that describe many NQF-endorsed measures. We would encourage a potential user of the measure to use a similar denominator threshold. We note there are other measures (e.g., 30-day stroke mortality) used by the Centers for Medicare and Medicaid Services (CMS) for accountability where public reporting is reserved for larger hospitals; smaller hospitals receive their (less precise) results as a quality improvement tool, rather than for public accountability. We envision that the same sort of procedure would occur for this measure once implemented.

Risk adjustment: The risk-adjustment approach used for this measure in unique in that it compares the same patient population at two different points in time. In short, it compares the patient's short-term risk of stroke (1-30d post-discharge) to their underlying baseline risk (91-360d post-discharge). As noted in the measure documentation, there are disparities in how well hospital EDs diagnosis strokes in different subgroups (women, younger patients, and people of color are more likely to experience a misdiagnosed stroke). It is these very disparities in diagnosis that our measure aims to highlight. Adjusting for clinical risk factors or social risk factors would result in these variations being adjusted away.

Sufficient variation: As discussed with the Scientific Methods Panel, our ability to distinguish "good" from "bad" performers on this measure is exclusively a function of the limited data set that we had available for testing the measure. The data set included only Medicare fee-for-service patients, which typically represents only about 20% of hospital ED discharges. In real-world applications, where more complete data sets are likely available, the ability to

distinguish “good” from “bad” will be substantially more precise. As can be seen in the data presented as part of our measure developer comments (reproduced below as Figure 1a/b), the true practice variation is substantial, with hospital performance ranging from 0 to over 150 per 10,000 discharges, with hundreds of hospitals having measured rates ranging from 20 to 200 per 10,000. These data reflect a 10-year window, so this level of precision or greater is what one would expect from a complete 100% ED sample (5x the 20% Medicare sample) from each hospital when using the proposed 3-year rolling window of analysis. This could be accomplished using HCUP data from states with linkable SEDD-SID records (now nearly half). In other words, this problem noted by the FAH is a problem related to data availability, not the measure itself.

Figure 4. *From Medicare data using the method proposed (Figure 4, shown as the measured 30d rate above expected)…*

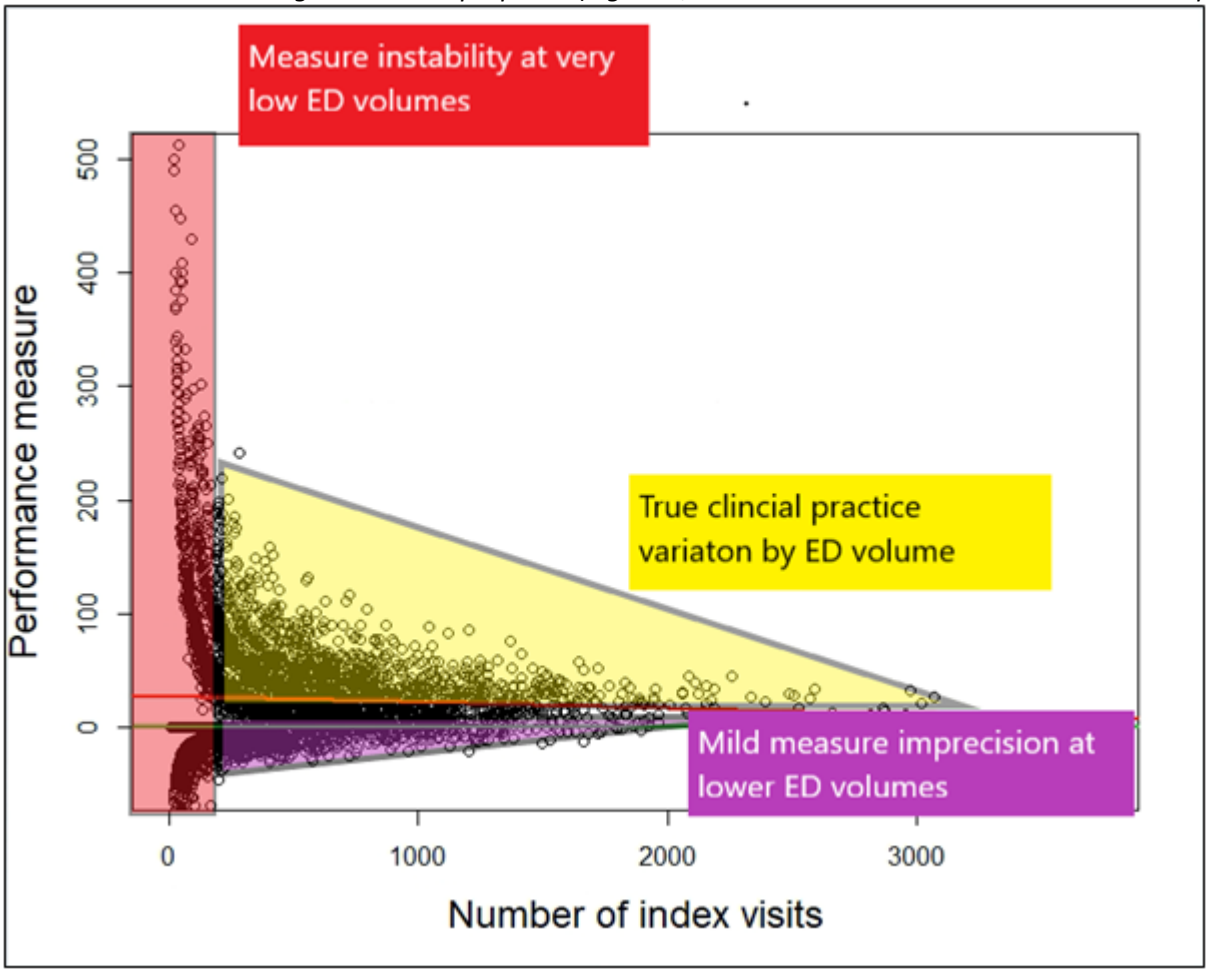


Figure 4. Excess short-term stroke rates at all hospitals by ED visit volume, with descriptive overlay separating true variation from measure imprecision. These Medicare data reflect 5,472 facilities over a 10-year window from 2009-2018. Each circle represents a single facility. The Figure demonstrates that smaller facilities have higher 30-day stroke hospitalization rates above the expected base rate after ED treat-and-release visit (TRV) for “benign dizziness.” Optimal measure performance is to have a zero rate above baseline (0 on the Y axis). The graph shows wide variation in ED performance on the measure (from less than zero to 500 excess stroke hospitalizations per 10,000 TRVs). Although not all of this variation reflects actual clinical performance, the vast majority of US hospitals have non-zero (>0) rates. The regression trend line shows the association between facility size and measure performance, with the larger facilities having the best performance (zero excess strokes over expected). The ***red shaded area*** reflects measure instability at the smallest hospitals. For hospitals with fewer than ~20 treat-and-release visits (TRV) for “benign dizziness” each year, the measure would be used only for quality improvement and **not** public accountability. The ***purple shaded area*** shows mild measure imprecision at hospitals with 20-200 dizziness TRVs each year; maximum imprecision is +/- ~20 per 10,000 TRVs at the

smaller EDs. The ***yellow shaded area*** shows true clinical performance variability (from rates of 0 excess strokes per 10,000 TRVs to >150 excess strokes – i.e., 1.5% of all "benign" discharges). ***This is strong evidence of wide practice variation around the US***.

NQF Response

Thank you for your comment.

NQF Committee Response

Thank you for your comment. The Standing Committee reviewed and discussed this information and the developer's response during the post-comment meeting and agreed with many of the concerns presented. The Standing Committee ultimately did not pass the measure on Evidence.

NQF #3614 Hospitalization After Release with Missed Dizzy Stroke (H.A.R.M Dizzy-Stroke), Comment #7835

Standing Committee Recommendation: Measure Not Recommended for Endorsement

Comment ID#: 7835

Commenter: Submitted by J. Matthew Austin

Council / Public: QMRI

Comment Period: Post-Evaluation Public and Member Commenting

Date Comment was Submitted: 9/27/2021

Developer Response Required? No

Level of Support: N/A

Theme: N/A

Comment

NQF Comment Period Additional Evidence for Measure #3614

Contents of this Document

Below the measure developers offer additional evidence to meet the NQF standard for "Evidence." Some members of the Neurology Standing Committee, in their initial review of Evidence, did not see a clear link between the measure, the quality improvements that would be induced, and the outcomes for patients. **Part I** below defines the logical links and accompanying evidence supporting the relationship between the measure and improved patient outcomes. **Part II** below shows how, in addition, the measure clearly meets the NQF standard for Evidence on purely technical grounds.

Measure #3614

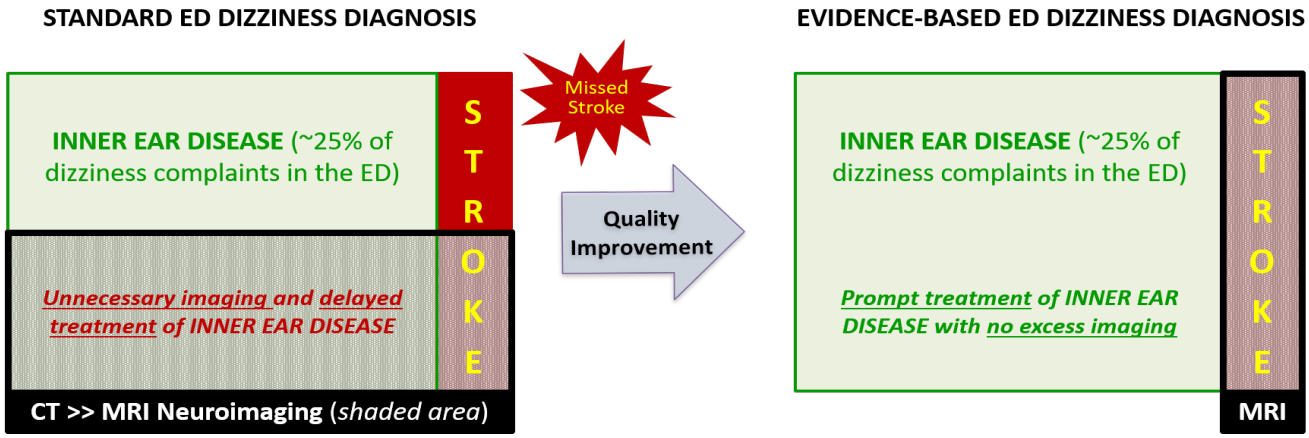
Avoid Hospitalization After Release with Missed Dizzy Stroke (Avoid H.A.R.M. Dizzy-Stroke)

The measure denominator is ED treat-and-release with "benign" dizziness. The measure numerator is observed 30-day stroke hospitalizations post ED treat-and-release with "benign" dizziness, minus the expected number of stroke hospitalizations occurring during that same period.

Part I. Logic Model and Supporting Evidence for Improved Quality and Patient Outcomes

A. Stepwise mechanism by which proposed measure will improve quality/safety for patients…

1. Measure #3614 reflects missed strokes in ED patients presenting with dizziness or vertigo
2. Accountability to the measure requires QI efforts that improve ED diagnosis of dizziness/vertigo (Figure 1)
3. These QI efforts will improve diagnosis both for patients with stroke and inner ear disease
4. Benefits to patients will then accrue from the prompt application of RCT-proven treatments
    a. Those with stroke will benefit from tPA or early secondary prevention, as appropriate
    b. Those with BPPV will benefit from prompt canalith repositioning and less CT radiation
5. These benefits to stroke patients (4a), in turn, will result in a "better" measure score (Figure 2)



CURRENT PRACTICE: Search for stroke is mostly based on non-selective imaging (~90% by CT) of dizziness based on patient age and vascular risk rather than exam. Estimated ~45,000-75,000 missed strokes annually, many causing harms.

NEW PRACTICE: Bedside exams lead to selective imaging (MRI) to confirm stroke, pinpoint cause, and guide stroke treatments. Estimated ~25,000 harms prevented and ~$1 billion in costs saved (half from unnecessary CTs, half admissions).

Figure 1. Theory for ED practice change. Standard practice in diagnosing dizziness now rests largely on CT to search for stroke in older patients with vascular risk factors. However, CT is ineffective for diagnosing vestibular strokes. Because inner ear causes are also more common among older populations with stroke risk factors, imaging is overused in inner ear diseases. Simultaneously, young patients (or old patients without vascular risk factors) who *do* have strokes as the cause may inadvertently be sent home untreated, sometimes with devastating consequences.[1,2] QI interventions such as teleconsultation will focus neuroimaging on directing stroke treatments, and more patients with inner ear disease will be correctly diagnosed and treated, preventing unnecessary imaging and admission.

Abbreviations: CT– computerized tomography; MRI – magnetic resonance imaging; QI – quality improvement

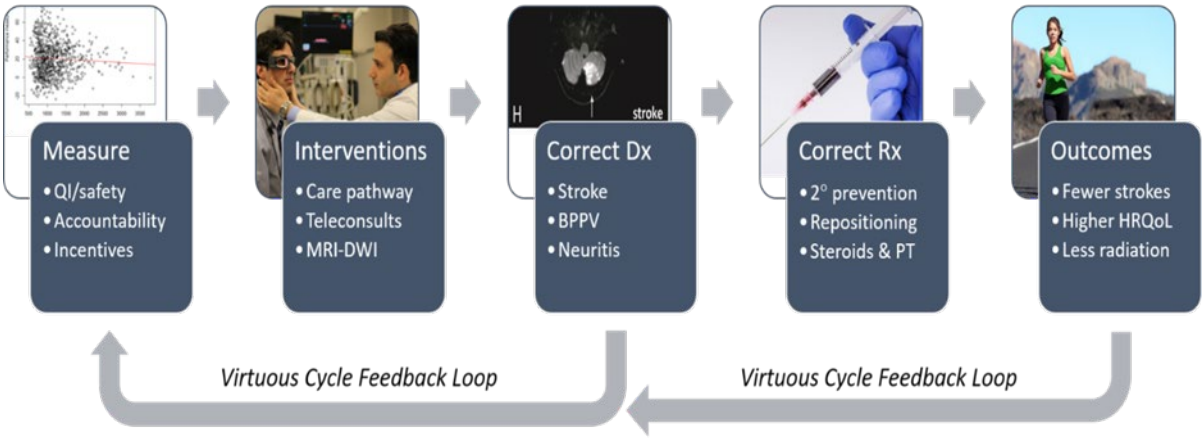


Figure 2. Logic model by which proposed measure will improve quality and safety for patients.

Abbreviations: 2° – secondary; Dx – diagnosis; HRQoL – health related quality of life; MRI-DWI – magnetic resonance imaging with diffusion weighted images; PT – physical therapy; QI – quality improvement; Rx – treatment

B. Logical validity of the evidence supporting positive impact of the measure on patient care…

1. **SYSTEMATIC REVIEW EVIDENCE THAT BETTER EYE EXAMS INCREASE CLINICAL DIAGNOSTIC ACCURACY**: There is strong evidence from multiple systematic reviews with meta-analyses of multiple prospective observational studies that bedside eye movement exams ("HINTS") in the hands of neurologists can more accurately diagnose stroke in patients with dizziness than even MRI scans.[3-6] Furthermore, the accuracy of these bedside exams far exceeds that of the more commonly used imaging technique of CT (which misses over 90% of acute posterior fossa strokes presenting with dizziness [reviewed in Newman-Toker, 2016[7]]), as well as the overall accuracy of current ED care, in which 40% of strokes presenting with dizziness are missed.[8] Neurology consultation services directly to the ED have demonstrated dramatically improved diagnostic accuracy, while simultaneously reducing inappropriate imaging.[9,10] Reductions in inappropriate CT use eliminate unnecessary irradiation, thereby cutting cancer risk, so improving outcomes for patients.[11] And while untrained ED clinicians do not perform this bedside testing well, those who are trained using direct observation and feedback methods achieve similar diagnostic results to those obtained by specialists—(sensitivity: 92.9% [95% CI 70–100%]; specificity: 96.4% [95% CI 93–98%]; positive predictive value: 81.3% [95% CI 61–87%]; negative predictive value: 98.8% [95% CI 95–100%]).[12]

2. **FACE VALIDITY THAT BETTER DIAGNOSIS YIELDS BETTER TREATMENT**: It is face valid that increasing correct diagnosis of posterior stroke in patients with dizziness and vertigo will lead to greater application of randomized trial and guideline approved stroke treatments in the ED. Likewise the same for inner ear diseases.

3. **RCT EVIDENCE THAT EARLY TREATMENT OF MINOR STROKE/TIA IMPROVES OUTCOMES**: It is proven through randomized clinical trials (CHANCE, POINT) that certain patients with TIA and minor stroke benefit from the application of early secondary prevention treatments, such as dual antiplatelet therapy. Combined results in over 10,000 patients show that treatment in the first 24 hours cuts the risk of a major stroke by 34% in the next 21 days.[13] In our original application for measure #3614, we provided similar empirical evidence from other studies of the benefit of immediate stroke treatments: "Preventable adverse outcomes of misdiagnosis result from missed opportunities for thrombolysis,[14,15] early surgery for malignant posterior fossa edema,[16,17] or prevention of subsequent infarction.[18-20] Rapid treatment improves health outcomes[21,22] and prompt prophylaxis lowers repeat stroke risk by up to 80%.[23,24] Thus, patients generally benefit from early, correct diagnosis."

4. **RCT EVIDENCE THAT EARLY TREATMENT OF INNER EAR DISEASES IMPROVES OUTCOMES**: Benefits also accrue to patients with dizziness or vertigo who are correctly diagnosed with inner ear disease (benign paroxysmal positional vertigo and vestibular neuritis) who receive guideline-supported treatments with randomized controlled trial evidence,[25-31] and direct harms of misdiagnosis[32] are reduced.

5. **FACE VALIDITY THAT PREVENTING MAJOR STROKES WILL LOWER THE MEASURE SCORE**: It is face valid that if there are fewer subsequent major strokes among those treated, then there will

be fewer short-term hospitalizations for stroke, which is, in turn, reflected in the measure (i.e., by reducing the "n" in the numerator). Furthermore, properly identifying such patients in the first place will remove these higher risk patients from the denominator (by correctly diagnosing stroke rather than "benign" inner ear disease or non-specific dizziness); this will tend to lower the observed number of subsequent strokes towards the expected population base rate (which is included as part of the measure calculation, which is observed minus expected).

C. Evidence of improved diagnostic accuracy in clinical practice with consult-based quality improvement…

Recent data (Table 1) from a quality improvement intervention (Tele-Dizzy) involving remote neurology consultations show dramatic *increases* in both stroke and specific inner ear diagnoses, along with dramatic *decreases* in inappropriate imaging among 287 patients who underwent consultation, relative to a matched emergency department population. These results provide compelling ***empirical evidence*** supporting the link between a ***healthcare intervention/service*** and the outcome of improved diagnosis, as well as better patient outcomes (reduction in unnecessary irradiation). It is inferentially logical and face valid, then, that these results, implemented more broadly, could be measured using #3614.

Table 1. Results of Tele-Dizzy Quality Improvement Intervention at Johns Hopkins Hospital (n=287 teleconsults).

Category	Parameter	Baseline*	Tele-Dizzy	Improvement	p-value (χ^2)
Diagnostic Yield	Specific Vestibular Diagnosis Rate	77 (20.6%)	163 (56.8%)	↑ 176%	P<0.0001
	Stroke Diagnosis Rate	1 (0.3%)	20 (7.0%)	↑ 2,506%	P<0.0001
	Non-Diagnosis Rate	235 (62.8%)	86 (30.0%)	↓ 52%	P<0.0001
Test Utilization	Neuroimaging (CT or MRI)	198 (52.9%)	70 (24.4%)	↓ 54%†	P<0.0001
Patient Outcomes	Excess 30-day stroke hospitalizations	0.1%‡	0 (0.0%)‡	↓ 100%‡	NA

** Baseline rates for diagnostic accuracy and test utilization are from 374 ED patients with a presenting symptom of dizziness (seen outside of Tele-Dizzy consultation hours) who had mention of "nystagmus" in notes and were comparable on the variables age, sex, and ED triage acuity.*

† CT scans were reduced by 96% (from 49.2% to 1.7%, p<0.0001) and MRIs for patients without strokes were unchanged (15.5% vs. 15.7%, p=0.95).

‡ Baseline 30d stroke hospitalizations are calculated as in Measure #3614 (not using the comparator population for Tele-Dizzy, which was too small for a precise estimate). The Tele-Dizzy value is based on actual patients seen at the same hospital – thus far, there have been zero stroke returns.

Part II. NQF Evidence Standard for Outcomes Measures (directly quoted from NQF documents, bold emphasis added)

"1a. Evidence. The evidence requirements for a health outcome measure include providing ***empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service***; if these data [are] not available, ***data demonstrating wide variation in performance***, assuming the data are from a robust number of providers and results are not subject to systematic bias."

Technical Elements of NQF Evidence Standard Met by Measure #3614

A. *"…empirical data that demonstrate a relationship between the outcome and at least one healthcare **structure**…"*

B. *"…empirical data that demonstrate a relationship between the outcome and at least one healthcare **process**…"*

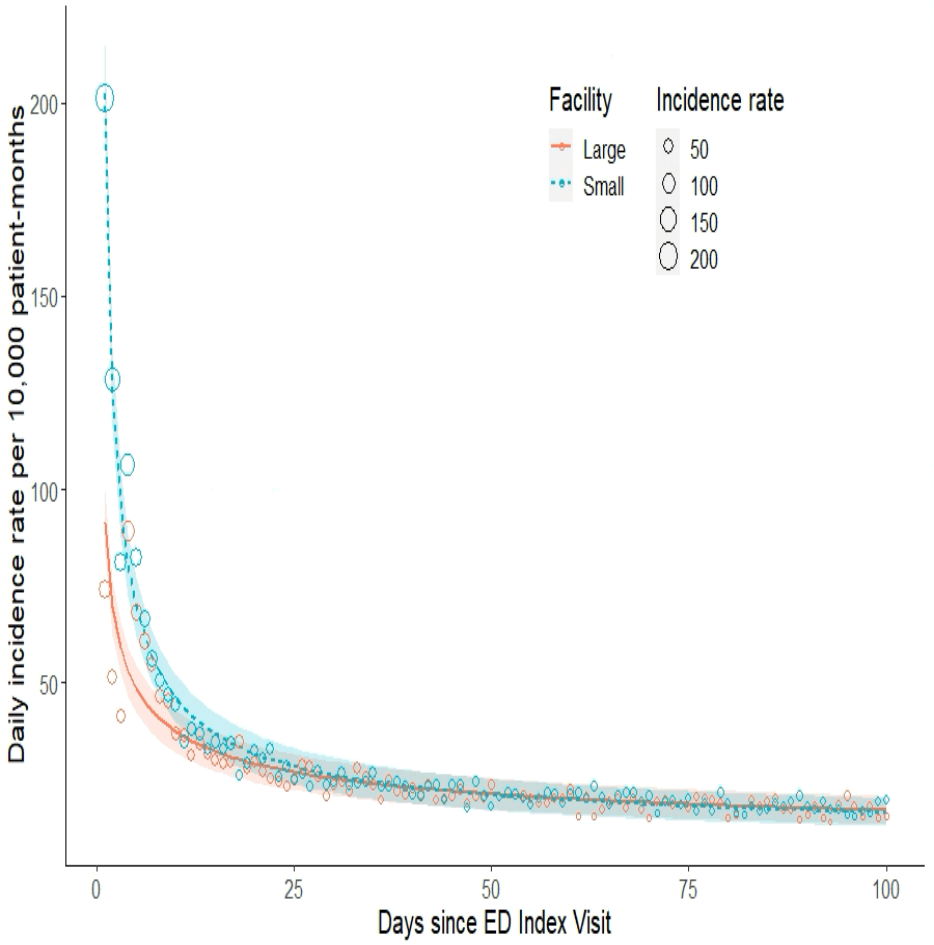
C. *"…data demonstrating wide **variation** in performance…"*

PAGE 59

*NOTE – Only **ONE** of the three options is needed to meet the NQF Evidence standard, but all three are met. Also, only **ONE** structure OR **ONE** process is required, but FOUR structures and THREE processes are empirically shown below.*

A. HEALTHCARE STRUCTURE (ED volume, population size, teaching status, % admitted from ED)

From Medicare data using the method proposed (Figure 3, shown as incidence rate curve) …



From Medicare data using the method proposed (Figure 3, shown as incidence rate curve) …

Figure 3. Short-term stroke rates at large vs. small hospitals (based on ED visit volumes). The graph at left shows stroke hospitalization incidence rates for the first 100 days after an ED treat-and-release visit (TRV) for "benign dizziness." Red represents larger hospitals and blue smaller hospitals. Shaded areas are 95% confidence intervals. These Medicare data reflect 5,472 facilities over a 10-year window from 2009-2018. A cutoff of 1,000 ED index visits over 10 years was used to define large vs. small facilities (1,472,612 ED TRVs occurred in large facilities, and 1,422,724 ED TRVs in small facilities). The Figure demonstrates that smaller facilities have higher short-term stroke incidence, mostly in the first 2 weeks after ED treat-and-release visit for "benign dizziness." These represent missed strokes in the ED.

From HCUP data (Newman-Toker, 2014[33]) using a similar method to that proposed…

Data element	Value	EST	SE	Z	P	OR	LCL	UCL
Hospital Characteristics								
Region								
	Midwest	-0.17	0.15	-1.1	0.27	0.84	0.62	1.14
	South	-0.12	0.13	-0.93	0.35	0.88	0.68	1.14
	West	-0.03	0.12	-0.24	0.81	0.97	0.77	1.22
	Northeast							
Population size								
	Small metropolitan	-0.26	0.09	-2.94	0.003	**0.77**	**0.65**	**0.92**
	Micropolitan	0.21	0.15	1.41	0.16	1.23	0.92	1.64
	Rural	0.06	0.24	0.24	0.81	1.06	0.67	1.68
	Large metropolitan							
Ownership								
	Public	-0.01	0.12	-0.10	0.92	0.99	0.78	1.25
	Private, for-profit	-0.22	0.12	-1.93	0.05	0.80	0.64	1.00
	Private, not-for-profit							
Teaching Status								
	Nonteaching	0.37	0.11	3.24	<0.001	**1.45**	**1.16**	**1.82**
	Teaching							
Hospital workflow (annual average)								
Inpatient occupancy rate (annual)								
	Low <=0.5	0.00	0.13	0.03	0.98	1.00	0.78	1.29
	Moderate >0.5, <0.7	0.11	0.11	0.94	0.35	1.11	0.89	1.39
	High >=0.7							
ED Volume (annual)								
	Low <=29,124	0.45	0.17	2.69	0.007	**1.57**	**1.13**	**2.18**
	Moderate 29-125-64,434	0.10	0.10	1.02	0.31	1.11	0.91	1.36
	High >=64,435							
Percent admitted from ED (annual)								
	Low <=11.82%	0.44	0.15	2.88	.004	**1.55**	**1.15**	**2.09**
	Moderate >11.82, <19.46%	0.21	0.11	1.95	0.05	1.24	1.00	1.54
	High >=19.46%							

B. HEALTHCARE PROCESS (weekend visit, ED admit rate on day of visit, patient left AMA)

From HCUP data (Newman-Toker, 2014[33]) using a similar method to that proposed…

Data element	Value	EST	SE	Z	P	OR	LCL	UCL
ED Visit characteristics (day of initial treat-and-release visit)								
Weekend								
	Monday-Friday	0.11	0.05	2.09	0.04	**1.11**	**1.01**	**1.23**
	Saturday-Sunday							
ED Crowding on day of visit (percentile)								
	0-20th percentile	-0.02	0.07	-0.33	0.75	0.98	0.84	1.13
	21-40th percentile	0.04	0.07	0.5	0.62	1.04	0.90	1.19
	41-60th percentile	0.04	0.07	0.52	0.60	1.04	0.90	1.20
	61-80th percentile	0.08	0.07	1.18	.0.24	1.08	0.95	1.23
	81-100th percentile							
ED admit rate on day of visit (percentile)								
	0-20th percentile	1.85	0.16	11.72	<0.001	**6.34**	**4.66**	**8.63**
	21-40th percentile	0.91	0.11	8.03	<0.001	2.48	1.99	3.10
	41-60th percentile	0.61	0.10	6.05	<0.001	1.85	1.51	2.25
	61-80th percentile	0.34	0.08	4.04	<0.001	1.40	1.19	1.66
	81-100th percentile							
Patient left against medical advice								
	Against medical advice	1.08	0.14	7.50	<0.001	**2.94**	**2.22**	**3.89**
	Not against advice							

187,188 of 198,819 trials used; number of events used =2088 of 2243 (records with missing data excluded); exchangeable correlation structure (working correlation = 0.002); 1016 clusters (facilities). EST, estimate; SE, standard error; Z, Z score; p, probability level; OR, odds ratio; LCL, lower confidence limit; UCL, upper confidence limit. [a]This is a patient-level analysis of inpatient stroke admissions, with and without a prior treat-and-release ED visit for dizziness or headache within 30 days of the stroke admission; only a single 'initial' ED visit (the most proximate to the 'index' stroke admission) is considered.

C. WIDE VARIATION IN PERFORMANCE

From Medicare data using the method proposed (Figure 4, shown as the measured 30d rate above expected)…

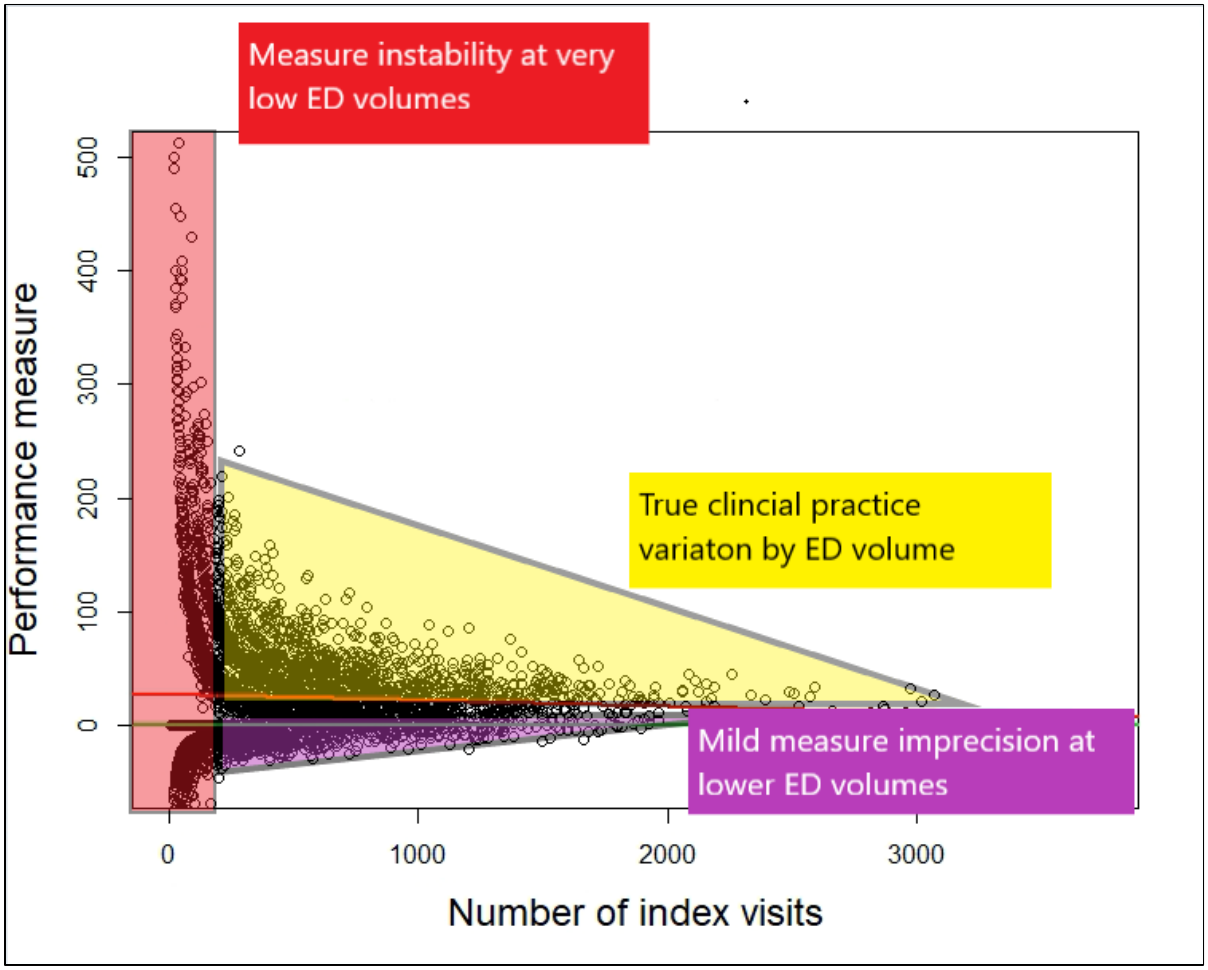


Figure 4. Excess short-term stroke rates at all hospitals by ED visit volume, with descriptive overlay separating true variation from measure imprecision. These Medicare data reflect 5,472 facilities over a 10-year window from 2009-2018. Each circle represents a single facility. The Figure demonstrates that smaller facilities have higher 30-day stroke hospitalization rates above the expected base rate after ED treat-and-release visit (TRV) for "benign dizziness." Optimal measure performance is to have a zero rate above baseline (0 on the Y axis). The graph shows wide variation in ED performance on the measure (from less than zero to 500 excess stroke hospitalizations per 10,000 TRVs). Although not all of this variation reflects actual clinical performance, the vast majority of US hospitals have non-zero (>0) rates. The regression trend line shows the association between facility size and measure performance, with the larger facilities having the best performance (zero excess strokes over expected). The ***red shaded area*** reflects measure instability at the smallest hospitals. For hospitals with fewer than ~20 treat-and-release visits (TRV) for "benign dizziness" each year, the measure would be used only for quality improvement and **not** public accountability. The ***purple shaded area*** shows mild measure imprecision at hospitals with 20-200 dizziness TRVs each year; maximum imprecision is +/- ~20 per 10,000 TRVs at the smaller EDs. The ***yellow shaded area*** shows true clinical performance variability (from rates of 0 excess strokes per 10,000 TRVs to >150 excess strokes – i.e., 1.5% of all "benign" discharges). ***This is strong evidence of wide practice variation around the US***.

SUMMARY

Part I above offers a set of valid logical links between the measure, quality improvement interventions, and improved patient outcomes. Each of the key steps is either supported by strong empirical evidence or is naturally face valid. Although none of this is required to meet the NQF "Evidence" standard for outcome measures, this, nevertheless, directly addresses the Neurology Standing Committee concerns regarding the strength of underlying evidence.

Part II above shows how the measure clearly meets the NQF "Evidence" standard on technical grounds. The measure must only demonstrate that it meets one such element, but we provide evidence that it meets the standard nine times.

Parts of the evidence presented here were submitted with the original NQF Evidence Attachment. This includes specific citations to randomized trials evidence of benefit to patients with early diagnosis and prompt treatments (quoted in Part I) and at least two components of the technical standard (Figure 3 and part of Figure 4). In accord with this, NQF staff, in their pre-review of the measure, concluded that the Evidence Criterion had been adequately passed by #3614.

Thus, in summary, we are confident that measure #3614 meets both the spirit and the letter of the standard.

Therefore, we hope that the Committee will reconsider its initial vote, and vote "pass" on the Evidence criterion.

References

1. Savitz SI, Caplan LR, Edlow JA. Pitfalls in the diagnosis of cerebellar infarction. Acad Emerg Med. 2007;14(1):63-8.
2. Missed Stroke Diagnosis - John Michael Night's Story. Society to Improve Diagnosis in Medicine; 2020 [updated 2020; cited 2020 February 12]; Available from: https://www.improvediagnosis.org/stories_posts/missed-stroke-diagnosis/.
3. Tarnutzer AA, Berkowitz AL, Robinson KA, Hsieh YH, Newman-Toker DE. Does my dizzy patient have a stroke? A systematic review of bedside diagnosis in acute vestibular syndrome. Canadian Medical Association Journal. 2011;183(9):E571-92. PMCID: 3114934.
4. Cohn B. Can Bedside Oculomotor (HINTS) Testing Differentiate Central From Peripheral Causes of Vertigo? Annals of emergency medicine. 2014.
5. Krishnan K, Bassilious K, Eriksen E, Bath PM, Sprigg N, Braekken SK, Ihle-Hansen H, Horn MA, Sandset EC. Posterior circulation stroke diagnosis using HINTS in patients presenting with acute vestibular syndrome: A systematic review. Eur Stroke J. 2019;4(3):233-9. PMCID: PMC6960692.
6. Ohle R, Montpellier RA, Marchadier V, Wharton A, McIsaac S, Anderson M, Savage D. Can Emergency Physicians Accurately Rule Out a Central Cause of Vertigo Using the HINTS Examination? A Systematic Review and Meta-analysis. Acad Emerg Med. 2020;27(9):887-96.
7. Newman-Toker DE, Della Santina CC, Blitz AM. Vertigo and hearing loss. Handb Clin Neurol. 2016;136:905-21.
8. Tarnutzer AA, Lee SH, Robinson KA, Wang Z, Edlow JA, Newman-Toker DE. ED misdiagnosis of cerebrovascular events in the era of modern neuroimaging: A meta-analysis. Neurology. 2017;88(15):1468-77. PMCID: PMC5386439.
9. Nham B, Reid N, Bein K, Bradshaw AP, McGarvie LA, Argaet EC, Young AS, Watson SR, Halmagyi GM, Black DA, Welgampola MS. Capturing vertigo in the emergency room: three tools to double the rate of diagnosis. J Neurol. 2021.
10. Gold D, Peterson S, McClenney A, Tourkevich R, Brune A, Choi W, Shemesh A, Maliszewski B, Bosley J, Otero-Millan J, Fanai M, Roberts D, Tevzadze N, Zee DS, Newman-Toker DE. Diagnostic impact

of a device-enabled remote "Tele-Dizzy" consultation service [abstract]. Diagnostic Error in Medicine, 12th Annual Conference; November 10-13, 2019; Washington, DC2019.
11. Smith-Bindman R, Lipson J, Marcus R, Kim KP, Mahesh M, Gould R, Berrington de Gonzalez A, Miglioretti DL. Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Archives of Internal Medicine. 2009;169(22):2078-86.
12. Vanni S, Nazerian P, Casati C, Moroni F, Risso M, Ottaviani M, Pecci R, Pepe G, Vannucchi P, Grifoni S. Can emergency physicians accurately and reliably assess acute vertigo in the emergency department? Emergency medicine Australasia : EMA. 2015.
13. Pan Y, Elm JJ, Li H, Easton JD, Wang Y, Farrant M, Meng X, Kim AS, Zhao X, Meurer WJ, Liu L, Dietrich D, Wang Y, Johnston SC. Outcomes Associated With Clopidogrel-Aspirin Use in Minor Stroke or Transient Ischemic Attack: A Pooled Analysis of Clopidogrel in High-Risk Patients With Acute Non-Disabling Cerebrovascular Events (CHANCE) and Platelet-Oriented Inhibition in New TIA and Minor Ischemic Stroke (POINT) Trials. JAMA Neurol. 2019;76(12):1466-73. PMCID: PMC6704730.
14. Kuruvilla A, Bhattacharya P, Rajamani K, Chaturvedi S. Factors associated with misdiagnosis of acute stroke in young adults. Journal of Stroke and Cerebrovascular Diseases. 2011;20(6):523-7.
15. Cano LM, Cardona P, Quesada H, Mora P, Rubio F. [Cerebellar infarction: prognosis and complications of vascular territories]. Neurologia. 2012;27(6):330-5.
16. Edlow JA, Newman-Toker DE, Savitz SI. Diagnosis and initial management of cerebellar infarction. Lancet Neurol. 2008;7(10):951-64.
17. 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. 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; a journal of cerebral circulation. 2013;44(3):870-947.
18. Flossmann E, Rothwell PM. Prognosis of vertebrobasilar transient ischaemic attack and minor stroke. Brain. 2003;126(Pt 9):1940-54.
19. Rothwell PM, Buchan A, Johnston SC. Recent advances in management of transient ischaemic attacks and minor ischaemic strokes. Lancet Neurol. 2006;5(4):323-31.
20. Paul NL, Simoni M, Rothwell PM. Transient isolated brainstem symptoms preceding posterior circulation stroke: a population-based study. Lancet Neurology. 2013;12(1):65-71.
21. Hacke W, Kaste M, Bluhmki E, Brozman M, Davalos A, Guidetti D, Larrue V, Lees KR, Medeghri Z, Machnig T, Schneider D, von Kummer R, Wahlgren N, Toni D. Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke. N Engl J Med. 2008;359(13):1317-29.
22. Lyden P. Thrombolytic therapy for acute stroke--not a moment to lose. N Engl J Med. 2008;359(13):1393-5.
23. Lavallee PC, Meseguer E, Abboud H, Cabrejo L, Olivot JM, Simon O, Mazighi M, Nifle C, Niclot P, Lapergue B, Klein IF, Brochet E, Steg PG, Leseche G, Labreuche J, Touboul PJ, Amarenco P. A transient ischaemic attack clinic with round-the-clock access (SOS-TIA): feasibility and effects. Lancet Neurol. 2007;6(11):953-60.
24. Rothwell PM, Giles MF, Chandratheva A, Marquardt L, Geraghty O, Redgrave JN, Lovelock CE, Binney LE, Bull LM, Cuthbertson FC, Welch SJ, Bosch S, Alexander FC, Silver LE, Gutnikov SA, Mehta Z. Effect of urgent treatment of transient ischaemic attack and minor stroke on early recurrent stroke (EXPRESS study): a prospective population-based sequential comparison. Lancet. 2007;370(9596):1432-42.
25. Bhattacharyya N, Gubbels SP, Schwartz SR, Edlow JA, El-Kashlan H, Fife T, Holmberg JM, Mahoney K, Hollingsworth DB, Roberts R, Seidman MD, Steiner RW, Do BT, Voelker CC, Waguespack RW, Corrigan MD. Clinical Practice Guideline: Benign Paroxysmal Positional Vertigo (Update). Otolaryngol Head Neck Surg. 2017;156(3_suppl):S1-S47.
26. Fife TD, Iverson DJ, Lempert T, Furman JM, Baloh RW, Tusa RJ, Hain TC, Herdman S, Morrow MJ, Gronseth GS. Practice parameter: therapies for benign paroxysmal positional vertigo (an evidence-based

review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology. 2008;70(22):2067-74.
27. Hilton MP, Pinder DK. The Epley (canalith repositioning) manoeuvre for benign paroxysmal positional vertigo. Cochrane Database of Systematic Reviews. 2014(12).
28. van den Broek EM, van der Zaag-Loonen HJ, Bruintjes TD. Systematic Review: Efficacy of Gufoni Maneuver for Treatment of Lateral Canal Benign Paroxysmal Positional Vertigo with Geotropic Nystagmus. Otolaryngol Head Neck Surg. 2014;150(6):933-8.
29. Lopez-Escamez JA, Gamiz MJ, Fernandez-Perez A, Gomez-Finana M, Sanchez-Canet I. Impact of treatment on health-related quality of life in patients with posterior canal benign paroxysmal positional vertigo. Otology & Neurotology. 2003;24(4):637-41.
30. Hillier S, McDonnell M. Is vestibular rehabilitation effective in improving dizziness and function after unilateral peripheral vestibular hypofunction? An abridged version of a Cochrane Review. Eur J Phys Rehabil Med. 2016;52(4):541-56.
31. Strupp M, Zingler VC, Arbusow V, Niklas D, Maag KP, Dieterich M, Bense S, Theil D, Jahn K, Brandt T. Methylprednisolone, valacyclovir, or the combination for vestibular neuritis. NEnglJMed. 2004;351(4):354-61.
32. Sacco AY, Self QR, Worswick EL, Couperus CJ, Kolli SS, Munoz SA, Carney JK, Repp AB. Patients' Perspectives of Diagnostic Error: A Qualitative Study. J Patient Saf. 2020.
33. Newman-Toker DE, Moy E, Valente E, Coffey R, Hines AL. Missed diagnosis of stroke in the emergency department: a cross-sectional analysis of a large population-based sample. Diagnosis (Berl). 2014;1(2):155-66. PMCID: PMC5361750.

NQF Response

N/A

NQF Committee Response

Thank you for your comment. The Standing Committee reviewed and discussed this measure in full during the post-comment meeting but did not find the evidence sufficient to pass the measure at this time.

National Quality Forum
1099 14th Street NW, Suite 500
Washington, DC 20005
https://www.qualityforum.org