NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 0507

Corresponding Measures:

De.2. Measure Title: Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports

Co.1.1. Measure Steward: American College of Radiology (ACR)

De.3. Brief Description of Measure: Percentage of final reports for carotid imaging studies (neck magnetic resonance angiography (MRA), neck computerized tomographic angiography (CTA), neck duplex ultrasound, carotid angiogram) performed that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement

1b.1. Developer Rationale: There is wide variation in the use of methods for stenosis calculation, which may also lead to variation in the appropriateness of carotid intervention. Since the degree of stenosis is an important element of the decision for carotid intervention, characterization of the degree of stenosis needs to be standardized. Requiring that stenosis calculation be based on a denominator of distal internal carotid diameter or, in the case of duplex ultrasound, velocity measurements that have been correlated to angiographic stenosis calculation based on distal internal carotid diameter, makes the measure applicable to both imaging and duplex studies.

S.4. 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

S.6. Denominator Statement: All final reports for carotid imaging studies (neck MRA, neck CTA, neck duplex ultrasound, carotid angiogram) performed

S.8. Denominator Exclusions: No Denominator Exclusions or Denominator Exceptions

De.1. Measure Type: Process

S.17. Data Source: Claims, Registry Data

S.20. Level of Analysis: Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Oct 28, 2008 **Most Recent Endorsement Date:** Sep 23, 2016

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? This measure is not included in a composite.

Preliminary Analysis: Maintenance of Endorsement

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence
Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a ***structure, process or intermediate outcome*** measure is that it is based on a systematic review (SR) and grading of the body of empirical evidence where the specific focus of the evidence matches what is being measured. For measures derived from patient report, evidence also should demonstrate that the target population values the measured process or structure and finds it meaningful.

The developer provides the following evidence for this measure:

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

Evidence Summary of prior review in 2017

- The developer described that proper measurement of carotid stenosis is essential to achieving the well-established favorable results of carotid interventions such as endarterectomy and stenting, which can reduce the recurrence of stroke in patients with high grade stenosis. Accurate measurement technique is supported by clinical practice guidelines which encourage the use of standardized criteria for carotid artery assessments across a variety of imaging modalities.
- The authors link use of such standardized criteria for carotid artery measurement which were used in pivotal trials demonstrating linkages between carotid endarterectomy and improved outcomes (NASCET and others).
- The developer argues that since these trials show positive outcomes, clinical application of the trial data requires that the intervention population be as close as possible to the inclusion criteria of the successful study. For this reason, the developer describes that measuring the carotid stenosis on noninvasive imaging studies in keeping with the NASCET methodology is beneficial.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**
☒ **The developer provided updated evidence for this measure:**

- Moderate and severe stenosis (50-90%) carotid artery stenosis affects approximately 10% of the general population by their 8th decade and causes approximately 10% of all strokes.

- The stroke risk associated with asymptomatic carotid stenosis (ACS) falls 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.
- Most guidelines indicated that CEA or CAS were not recommended for mild ACS (<50%–70% by NASCET) or SCS (<50% by NASCET) by not including procedural recommendations or explicitly stating that these procedures should not be done or that medical treatment alone was indicated.

Exception to evidence

Based on staff review, there is not a clear link between outcomes/quality of care with accurate vs. inaccurate carotid measurement. If the Standing Committee agrees with staff, there is an option to move the measure forward with an exception if the Standing Committee agrees that it is ok (or beneficial) to hold providers accountable for performance in the absence of empirical evidence of benefits to patients.

Questions for the Standing Committee:

- *In 2017, the Standing Committee passed the measure for evidence with a moderate rating. In staff review of the evidence presented in this submission, there are concerns the evidence presented does not directly link stenosis measurement precision to improved outcomes. Rather the evidence presented suggests that carotid procedures are effective in reducing stroke risk. Does the Standing Committee believe the evidence presented is directly applicable to patient outcomes?*

Guidance from the Evidence Algorithm

Process measure based on systematic review with no direct evidence of the measure (Box 3) -> No empirical evidence is submitted (Box 7) -> There are performance measures of a related health outcome (Box 10) -> Insufficient

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

RATIONALE: There is not a clear link between outcomes/quality of care with accurate vs. inaccurate carotid measurement. Data provided shows carotid interventions are effective in reducing stroke risk and in those trials accurate measurement was performed. But no direct evidence is provided showing that accurate measurement itself or inaccurate measurement is associated with differences in outcomes.

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

Maintenance measures – increased emphasis on gap and variation

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

Gap data

- The developer provided the following registry data for measure performance gap:
    - 2012: Performance Rate: 16.85, # of patients included: 726555, # of physicians: 3186920, Min: 0.54, Max: 100, Interquartile range: 50

- 2013: Performance Rate: 24.91, # of patients included: 769239, # of physicians: 54732, Min: 0.63, Max: 100, Interquartile range: 31.25
- 2014: Performance Rate: 81.57, # of patients included: 772456, # of physicians: 18141, Min: 0.69, Max: 100, Interquartile range: 19.35
- 2015: Performance Rate: 75.40, # of patients included: 893220, # of physicians: 13417, Min: 0.60, Max: 100, Interquartile range: 16.7, Std Deviation: 20.84
- 2016: Performance Rate: 80.16, # of patients included: 1038759, # of physicians: 15779, Min: 2.11, Max: 100, Interquartile range: 7.83, Std Deviation: 16.93
- 2017: Performance Rate: 74.49, # of patients included: 686055, # of physicians: 11387, Min: 1.27, Max: 100, Interquartile range: 4, Std Deviation: 15.14
- 2018: Performance Rate: 74.97, # of patients included: 639413, # of physicians: 9129, Min: 0.06, Max: 100, Interquartile range: 1.11, Std Deviation: 11.93

Disparities

- In the current submission, the developer did not provide disparities data for whether the specific approach to carotid measurement differed by racial, age, or socioeconomic status.
- The developer did summarize literature addressing disparities whether patients received carotid stenosis testing. One study, Cheng et al (2012). found no racial disparity in receipt of carotid artery imaging detected within nonminority serving hospitals. However, the predicted probability of receiving carotid artery imaging for white patients at nonminority serving hospitals (89.7%, 95% CI [87.3%, 92.1%]) was significantly higher than both white patients (78.0% [68.3%, 87.8%] and black patients (70.5% [59.3%, 81.6%]) at minority serving hospitals.
- Another study found that black ischemic stroke patients were less likely to receive diagnostic carotid imaging than white patients, which was significant after the developer adjusted for risk.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure? Staff review suggests that this measure may have little room for improvement.
- Are you aware of evidence that disparities exist in whether specific approaches are used by radiologists in carotid imaging?

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

RATIONALE: Measurement has improved over the years and the IQR is a little over 1%.

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure."

- There is insufficient evidence that this process measure is related to outcome.

- Agree with staff there is no data to show that non-standardized carotid artery measurements fail to achieve outcomes.
- yes
- Always more difficult to evaluate the impact of a diagnostic test on quality and outcome because of the indirect relationship between the two, however I believe that this imaging measure improve radiology quality
- As noted it is not clear if the measure would affect the clinical outcome of interest - acute ischemic stroke
- Evidence directly relates to process being measured, and strong support exists in the literature for the targeted intervention to lead to desirable outcomes.

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care? Performance rates have been relatively stable, 75-82, from 2014-2018. Why is there no data from 2019 or 2020?

- Performance seems to be improving. The disparities provided appear to be more related to whether a patient receives testing than the measurement or lack thereof.
- yes
- This measure has had a positive impact on the practice of radiologists and remove it may bring us back.
- Some subgrouping was done but only on the performance of the test and not on the suggested measure
- Some evidence of disparities in administration of targeted intervention, lending support to the use of the proposed measure
- Performance seems to be improving. The disparities provided appear to be more related to whether a patient receives testing than the measurement or lack thereof.
- yes

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

2c. For composite measures: empirical analysis support composite approach

Reliability

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

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

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

Composite measures only:

2d. Empirical analysis to support composite construction. Empirical analysis should demonstrate that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct.

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

Evaluators: NQF Staff Review

Reliability

- The developer used a beta-binominal model to assess the signal-to-noise ratio at the performance score level across claims and registry data and reported individual reliability scores in Table 3 of the submission. The overall mean reliability score was 0.9340. The mean (CI) reliability is 0.99340 (0.99331, 0.99350), which was largely similar for each year reported 2015 to 2018.

Validity

- In 2020, the measure developer attempted to perform construct validity testing by correlating the results of #0507 to with other measures. However, they were unable to find suitable measures for this purpose within the same accountability program (MIPS).
- The developer then tried to demonstrate criterion validity using measure performance data at the population level. The plan was to perform the requisite analyses among these measures to determine if a relationship exists to support empirical validity, hypothesizing that hospitals or physicians performing well on several related measures (MIPS #409, MIPS #413 and HOQR OP-23/NQF #0661) would perform the same on the stenosis measure (NQF #0507). However, the developer was unable to format the measures' data sets to perform empirical analysis. While MIPS #409 and #413 were specified at the individual clinician level, CMS was unable to provide the developer with individual level data, because all submissions were done at the group level.
- Due to the issues above, the developer performed a new face validity study in November 2020. Note the face validity was also separately measured in the 2015 submission.
- The results of the 2020 face validity results demonstrated that 82.15% (23 members) of the panel either strongly agreed or agreed that this measure accurately distinguishes good from poor quality. Two panel members disagreed that the measure would accurately distinguish good from poor quality. One member stated that literature shows that CTA underestimates the stenosis and MRA overestimates the stenosis, compared to NASCET. The other member stated that ultrasound should be removed from the measure. He also added that stenosis on ultrasound is measured using velocities and NASCET should never be applied to ultrasound.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?

Questions for the Committee regarding validity:

- The staff has judged the validity testing to be insufficient because empirical validity testing is required at measurement maintenance. Solely face validity testing was conducted by the developer. Is the developer's rationale for not performing empirical validity testing sufficient? Are there other measures/outcomes the developer could have used/considered for empirical validity testing?

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

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

RATIONALE: Empirical validity testing is expected during maintenance review unless the developer can provide a strong rational for not conducting empirical validity testing. The Standing Committee determines the strength of the rationale provided.

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

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- No concerns regarding reliability
- It seems reliable
- unclear some concern
- Good evidence for reliability
- I am concerned about the duplex aspect of the measure being consistently implemented and determined.
- No concerns

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- No
- No concerns
- unclear
- No
- See above, I will review the duplex component in more detail later.
- No concerns about reliability

2b1. Validity -Testing: Do you have any concerns with the testing results?

- Insufficient validity testing
- It seems that neither the planned nor executed studies really support validity.
- no
- Same comment again: re relationship between diagnostic imaging and quality/outcome
- I am concerned again about the inclusion of ultrasound in the measure.
- No concerns about validity

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment) 2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure? 2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- NA
- no concerns
- no
- No concern
- I did not find enough information in the proposal to determine issues of threats to validity.
- No risk adjustments

2b4-6. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- NA
- Some concern about the ultrasound issue
- no
- No concern
- For 2b5, each tst has its own issue (overcalling, undercalling the degree of stenosis) so not sure how this will be factored in. If the measure is to perform the calculation without concern for accuracy of the calculation then this will not be an issue.
- It is unclear if there is an advantage to one method of measuring distal internal carotid diameter, i.e. direct or indirect.

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

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

   Feasibility

   - This measure's data elements are abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
   - Some data elements are in defined fields in electronic clinical data.
   - The data elements are manually abstracted from the radiology report. The ACR is working to enable extraction of free text from radiology reports using Artificial Intelligence (AI) and Natural Language Processing (NLP).
   - The developer has indicated there are subscription fees to radiology groups for use of this measure.

Questions for the Committee:

- Is use of registry data through ACR a feasible way to measure quality in radiology groups?

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?

- No concerns
- Manual abstraction is costly and prone to errors.
- Some Concern
- Very feasible. Not onerous in clinical practice.
- These data elements should be available albeit in free text form in most circumstances.
- "Some data elements are in defined fields in electronic sources"
- No concerns
- Man

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

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

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

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

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

OR

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

Accountability program details

- Accountability Program: This measure is being used by the CMS Payment Program for accountability and reimbursement with 10,000 physicians and 2.4 million patients being included in the program for the measure.

- Public Reporting: This measure is also used for quality improvement within the ACR registries.

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

Feedback on the measure by those being measured or others

- The developer receives feedback on this measure through the claims and registry data using the CMS Quality Payment Program for MIPS.
- The developer reports feedback using Qualified Clinical Data Registries (QCDRs) where the users upload the measure data to the QCDR. The second method of reporting is through CMS' annual MIPS Feedback Reports. The feedback reports, aggregated at a high-level, are also based on CMS performance benchmarks (calculated in deciles).

Additional Feedback:

- The measure developer did not give additional feedback at this time.

***Questions for the Committee**:*

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

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

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

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

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

Improvement results

- According to the developer, performance results on this measure presented in 1b under Performance Gap indicate that the data has shown significant improvements.

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

- The developer notes that this measure has created more standardization for carotid imaging results while supporting increased communications between radiologists and referring physicians.

Potential harms

- The developer does not note any potential harms from implementation of this measure.

***Questions for the Standing Committee**:*

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

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

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

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided ?4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- No concerns
- No concerns
- yes
- High use

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations? 4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- No concerns
- No concerns
- yes
- High usability
- As before linking the adherence to the measure to improving stroke outcome is not clear. I see no potential harm.
- Evidence supports that carotid endarterectomy reduces the risk of recurrent stroke, especially with more significant stenosis. Critically important to accurately measure degree of stenosis for clinicians and patients to choose surgical intervention.

Criterion 5: Related and Competing Measures

Related or competing measures

- There are no related or completing measures for this measure.

Harmonization

- No harmonization is needed for this measure.

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

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- No concerns
- yes
- Not to my knowledge
- None that I am aware of
- No

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 6/10/2021

- **Of the zero NQF members who have submitted a support/non-support choice**

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 0507

Measure Title: Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports

Measure is:

☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   Submission document: "MIF_0507" document, items S.1-S.22

   NOTE*: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

RELIABILITY: TESTING

Type of measure:

☐ **Outcome (including PRO-PM)** ☐ **Intermediate Clinical Outcome** ☒ **Process**

☐ **Structure** ☐ **Composite** ☐ **Cost/Resource Use** ☐ **Efficiency**

Data Source:

☐ **Abstracted from Paper Records** ☒ **Claims** ☒ **Registry**
☐ **Abstracted from Electronic Health Record (EHR)** ☐ **eMeasure (HQMF) implemented in EHRs** ☐ **Instrument-Based Data** ☐ **Enrollment Data** ☐ **Other (please specify)**

Level of Analysis:

☐ **Individual Clinician** ☐ **Group/Practice** ☐ **Hospital/Facility/Agency** ☐ **Health Plan**
☐ **Population: Regional, State, Community, County or City** ☐ **Accountable Care Organization**
☐ **Integrated Delivery System** ☐ **Other (please specify)**

Submission document: "MIF_0507" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of **patient-level data** conducted?

   ☐ **Yes** ☒ **No**

6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

7. **Assess the results of reliability testing**

Submission document: Testing attachment, section 2a2.3

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☐ **No**

   ☐ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and **all** testing results):

    ☒ **High** (NOTE: Can be HIGH **only if** score-level testing has been conducted)

    ☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has **not** been conducted)

    ☐ **Low** (NOTE: Should rate **LOW** if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: Should rate **INSUFFICIENT** if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

VALIDITY: TESTING

12. **Validity testing level: ☒ Measure score ☐ Data element ☐ Both**

13. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

    Submission document: *Testing attachment, section 2b1.*

    ☐ **Yes**

    ☐ **No**

    ☒ **Not applicable** (data element testing was not performed)

14. **Method of establishing validity of the measure score:**

    ☒ **Face validity**

    ☐ **Empirical validity testing of the measure score**

    ☐ **N/A (score-level testing not conducted)**

15. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    Submission document: Testing attachment, section 2b1.

    ☐ **Yes**

    ☐ **No**

    ☐ **Not applicable** (score-level testing was not performed)

16. **Assess the method(s) for establishing validity**

    Submission document: **Testing attachment, section 2b2.2**

17. **Assess the results(s) for establishing validity**

    Submission document: **Testing attachment, section 2b2.3**

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

18. **Please describe any concerns you have with measure exclusions.**

    Submission document: Testing attachment, section 2b2.

19. **Risk Adjustment**

    Submission Document: Testing attachment, section 2b3

    19a. **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

    19b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

    ☐ Yes ☐ No ☐ Not applicable

    19c. **Social risk adjustment:**

    19c.1 Are social risk factors included in risk model? ☐ Yes ☒ No ☐ Not applicable

    19c.2 Conceptual rationale for social risk factors included? ☐ Yes ☒ No

    19c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☒ No

    19d. **Risk adjustment summary:**

    19d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

    19d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

    19d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

    19d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

    19d.5. Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

    19e**. Assess the risk-adjustment approach**

20. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

    Submission document: Testing attachment, section 2b4.

21. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

    Submission document: Testing attachment, section 2b5.

22. **Please describe any concerns you have regarding missing data.**

    Submission document: Testing attachment, section 2b6.

For cost/resource use measures ONLY:

23. **Are the specifications in alignment with the stated measure intent?**

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

24. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

25. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

☐ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☐ **Low** (NOTE: Should rate LOW if you believe that there **are** threats to validity and/or relevant threats to validity were **not assessed OR** if testing methods/results are not adequate)

☒ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level **is required**; if not conducted, should rate as INSUFFICIENT.)

26. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

ADDITIONAL RECOMMENDATIONS

29. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

Developer Submission

NQF #: 0507

Corresponding Measures:

De.2. Measure Title: Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports

Co.1.1. Measure Steward: American College of Radiology (ACR)

De.3. Brief Description of Measure: Percentage of final reports for carotid imaging studies (neck magnetic resonance angiography (MRA), neck computerized tomographic angiography (CTA), neck duplex ultrasound, carotid angiogram) performed that include direct or indirect reference to measurements of distal internal carotid diameter as the denominator for stenosis measurement

1b.1. Developer Rationale: There is wide variation in the use of methods for stenosis calculation, which may also lead to variation in the appropriateness of carotid intervention. Since the degree of stenosis is an important element of the decision for carotid intervention, characterization of the degree of stenosis needs to be standardized. Requiring that stenosis calculation be based on a denominator of distal internal carotid diameter or, in the case of duplex ultrasound, velocity measurements that have been correlated to angiographic stenosis calculation based on distal internal carotid diameter, makes the measure applicable to both imaging and duplex studies.

S.4. 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

S.6. Denominator Statement: All final reports for carotid imaging studies (neck MRA, neck CTA, neck duplex ultrasound, carotid angiogram) performed

S.8. Denominator Exclusions: No Denominator Exclusions or Denominator Exceptions

De.1. Measure Type: Process

S.17. Data Source: Claims, Registry Data

S.20. Level of Analysis: Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Oct 28, 2008 **Most Recent Endorsement Date:** Sep 23, 2016

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? This measure is not included in a composite.

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

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

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

2021_NQF_Evidence_Attachment_195_8.docx,0507_Evidence_MSF5.0_Data_2012_Final_Submission.doc

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

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the

new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

No

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 0507

Measure Title: Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 4/2/2021

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

Outcome

☐ Outcome:

☐ Patient-reported outcome (PRO):

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)

☐ Intermediate clinical outcome (*e.g., lab value*):

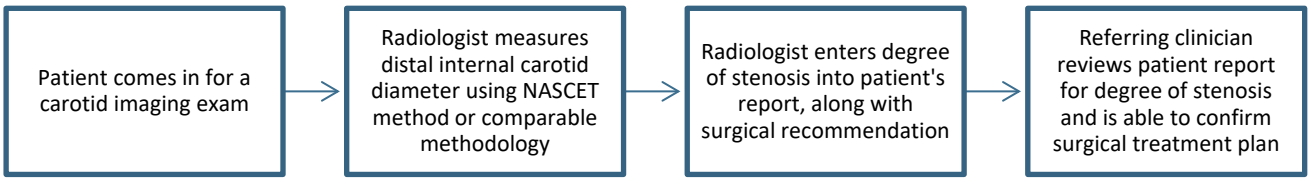
☒ Process:

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

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



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

Accurate assessment of the degree of carotid artery stenosis is essential to guiding proper treatment decisions for patients with carotid artery disease. Trials have demonstrated the ability of the degree of carotid artery

stenosis to predict which patients will receive the greatest benefit from surgical intervention. To ensure accurate assessment of stenosis, it is important to use a standardized, validated approach. A more accurate quantification of stenoses will lead to more appropriate treatment, based on the percentage of stenoses.

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

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

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Systematic Review	Evidence
Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Carotid endarterectomy for symptomatic carotid stenosis (Review) Orrapin S, Rekasem K 2017 Orrapin S, Rerkasem K. Carotid endarterectomy for symptomatic carotid stenosis. *Cochrane Database of Systematic Reviews* 2017, Issue 6. Art. No.: CD001081. DOI: 10.1002/14651858.CD001081.pub3.

Systematic Review	Evidence
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	With the exception of near-occlusions, the degree of stenosis above which surgery is beneficial was shown to be 50% (by the measurement used in the NASCET 1991 and VACSP 1991 trials: equivalent to about 65% stenosis by the method used in ECST 1998). Benefit in people with 50% to 69% stenosis became more modest with longer duration of follow-up. Lack of benefit for moderate stenosis in the original ECST 1998 report is not inconsistent with this but reflects the differences between the analyses in the measurement of stenosis and the definition of outcome events. The re-analysis of individual patient data has shown that the effects of surgery in ECST 1998 and NASCET 1991 in people with 50% to 69% stenosis were consistent. The process of standardizing the method for stenosis calculation, as indicated in the measure language, will lead to improved health outcomes such as more accurate quantification of stenoses and more appropriate treatment, based on the percentage of stenoses.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**Moderate** using GRADE scale. Moderate: The authors believe that the true effect is probably close to the estimated effect.
Provide all other grades and definitions from the evidence grading system	**Very low:** The true effect is probably markedly different from the estimated effect **Low:** The true effect might be markedly different from the estimated effect **Moderate:** The authors believe that the true effect is probably close to the estimated effect **High:** The authors have a lot of confidence that the true effect is similar to the estimated effect
Grade assigned to the **recommendation** with definition of the grade	**Moderate** using GRADE scale. **Moderate:** The authors believe that the true effect is probably close to the estimated effect.
Provide all other grades and definitions from the recommendation grading system	**Very low:** The true effect is probably markedly different from the estimated effect

Systematic Review	Evidence
	Low: The true effect might be markedly different from the estimated effect **Moderate:** The authors believe that the true effect is probably close to the estimated effect **High:** The authors have a lot of confidence that the true effect is similar to the estimated effect
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	This review identified three randomized controlled trials (6343 participants randomized), which compared carotid surgery with no carotid surgery (i.e. best medical therapy plus surgery versus best medical therapy alone) in participants with carotid stenosis and recent transient ischemic attacks (TIA) or minor ischemic strokes in the territory of that artery. The trials were carried out in Europe, USA, and Canada and included some centers in Israel, South Africa, and Australia. The gender ratio of participants was 2.6:1 (72% men and 28% women); 90% of participants were younger than 75 years old. Generally, the three included trials had adequate strategies to avoid bias in their study except VACSP 1991, which did not provide information on allocation concealment. Analysis of individual patient data has advantages over meta-analysis of overall trial results and was essential for the endarterectomy trials. Differences between the trials in the method of measurement of carotid stenosis and in the definition of outcome events made it impossible to combine tabular results satisfactorily. By re-analyzing the individual patient data and reassessing the carotid angiogram, the authors found that the results of ECST 1998 and NASCET 1991 were consistent, removing the uncertainty that was generated by the apparent disparities between the originally reported results of the trials.
Estimates of benefit and consistency across studies	Endarterectomy was of some benefit for participants with 50% to 69% symptomatic stenosis (moderate-quality evidence), and highly beneficial for those with 70% to 99% stenosis without near-occlusion (moderate-quality evidence). The authors found no benefit in people with carotid near-occlusion (high-quality evidence).

Systematic Review	Evidence
	The quality of the evidence for near occlusion and less than 30% of carotid stenosis is high. The quality of the evidence for 50% to 99% of carotid stenosis is moderate for any stroke or operative death as well as ipsilateral ischemic stroke and any operative stroke or death outcome. Patients with stenoses will benefit from physicians using a standardized method for stenosis calculation. Accuracy is extremely important as the calculation will justify the intervention selected for the patient, as evidence-based guidelines base treatment recommendations on the patient´s percentage of stenosis.
What harms were identified?	It is possible that the intention-to-treat analysis may have underestimated the benefit of endarterectomy in the near occlusions because of the relatively high rate of endarterectomy during follow-up in the medical treatment group in NASCET 1991. 70% to 99% stenosis without near-occlusion was significant for each of the three main outcomes. Some people may still wish to undergo surgery, particularly if they experience recurrent TIAs, but they should be informed that the benefit from endarterectomy in preventing a stroke is likely to be modest in the short-term and unknown in the long-term.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	The authors updated this review in September 2020. The results are still the same -- carotid endarterectomy reduced the risk of recurrent stroke for people with significant stenosis. Endarterectomy might be of some benefit for participants with 50% to 69% symptomatic stenosis (moderate-quality evidence) and highly beneficial for those with 70% to 99% stenosis (moderate-quality evidence).

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Systematic Review	Evidence
Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Systematic Review of Guidelines for the Management of Asymptomatic and Symptomatic Carotid Stenosis Anne L. Abbott, PhD, MBBS, FRACP; Kosmas I. Paraskevas, MD, PhD; Stavros K. Kakkos, MD, PhD; Jonathan Golledge, MB, BChir, BA, MA, MChir; Hans-Henning Eckstein, MD, PhD; Larry J. Diaz-Sandoval, MD; Longxing Cao, MD, PhD; Qiang Fu, MD, PhD; Tissa Wijeratne, MD, FRACP; Thomas W. Leung, MD; Miguel Montero-Baker, MD; Byung-Chul Lee, MD, PhD; Sabine Pircher, BNutrDiet, MPH; Marije Bosch, PhD; Martine Dennekamp, PhD, MSc; Peter Ringleb, MD, PhD August 31, 2015. Abbott, A. L., Paraskevas, K. I., Kakkos, S. K., Golledge, J., Eckstein, H. H., Diaz-Sandoval, L. J., Cao, L., Fu, Q., Wijeratne, T., Leung, T. W., Montero-Baker, M., Lee, B. C., Pircher, S., Bosch, M., Dennekamp, M., & Ringleb, P. (2015). Systematic Review of Guidelines for the Management of Asymptomatic and Symptomatic Carotid Stenosis. Stroke, 46(11), 3288–3301. https://doi.org/10.1161/STROKEAHA.115.003390 https://pubmed.ncbi.nlm.nih.gov/26451020/
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Moderate and severe (50%–99%) carotid artery stenosis is an important public health issue. This condition affects ≈10% of the general population by their 8th decade, and it causes ≈10% of all strokes. For many years, procedural management has been commonly recommended for stroke prevention. However, important relatively recent discoveries should improve treatment decisions for patients with carotid stenosis. These include: 1. The 60% to 80% fall in stroke risk associated with asymptomatic carotid stenosis (ACS) with medical treatment alone (encouraging a healthy lifestyle

Systematic Review	Evidence
	and appropriate medication) since the start of the randomized trials of 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) given medical treatment, with or without additional CEA. 2. Stroke risk stratification studies of patients with ACS showing that transcranial embolus detection, degree of stenosis, plaque echolucency, and asymptomatic progression are not sufficiently powerful individually to identify asymptomatic patients likely to benefit from carotid procedures. Combinations of markers are most likely to provide clinically meaningful stroke risk stratification. 3. Falls in the risk of stroke or death associated with CEA for patients with ACS or SCS. 4. The significantly higher overall risk of stroke or death associated with carotid angioplasty/stenting (CAS) than with CEA.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**Moderate** using GRADE scale. Moderate: The authors believe that the true effect is probably close to the estimated effect.
Provide all other grades and definitions from the evidence grading system	**Very low:** The true effect is probably markedly different from the estimated effect **Low:** The true effect might be markedly different from the estimated effect **Moderate:** The authors believe that the true effect is probably close to the estimated effect **High:** The authors have a lot of confidence that the true effect is similar to the estimated effect
Grade assigned to the **recommendation** with definition of the grade	**Moderate** using GRADE scale. **Moderate:** The authors believe that the true effect is probably close to the estimated effect.
Provide all other grades and definitions from the recommendation grading system	**Very low:** The true effect is probably markedly different from the estimated effect

Systematic Review	Evidence
	Low: The true effect might be markedly different from the estimated effect **Moderate:** The authors believe that the true effect is probably close to the estimated effect **High:** The authors have a lot of confidence that the true effect is similar to the estimated effect
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Each guideline was checked for completeness in defining asymptomatic carotid stenosis (ACS) and symptomatic carotid stenosis (SCS) within the target populations for the degree of stenosis, method of quantifying stenosis (North American Symptomatic Carotid Endarterectomy Trial [NASCET], European Carotid Surgery Trial [ECST], or other), and the timing and territory/laterality of any previous clinically defined strokes or transient ischemic attacks (TIA). The authors included all guideline recommendations for routine practice use of CEA and CAS published from January 1, 2008, to January 28, 2015. To be considered a guideline, it had to include a recommendation covering carotid endarterectomy (CEA) and/or carotid artery angioplasty/stenting (CAS) and/or SCS, or both based on evidence. The authors identified 34 guidelines meeting the inclusion criteria. These were sets of recommendations on CEA or CAS, or both for ACS or SCS, or both published between January 1, 2008, and January 28, 2015, in 41 separate documents from 23 different regions/countries (including 2 representing Europe and 5 the United States). They were written by 32 different groups in 6 languages (English, Chinese, Korean, Spanish, Dutch, and German). One group (American Heart Association/ American Stroke Association) published a guideline on carotid stenosis for men and women together and a separate one for women only; both were included in this study.
Estimates of benefit and consistency across studies	Only 2 of 28 (7%) guidelines with procedural recommendations on ACS completely defined ACS according to degree of stenosis, method of determining degree of stenosis, and timing and territory of any previous stroke or TIA. Even then, in 1 case, the timing

Systematic Review	Evidence
	of any previous stroke or TIA (<6 months) was deduced from the definition of SCS. Three guidelines contained no definition of ACS. Among the remaining 23 guidelines, degree of stenosis was always specified, and 1 distinct cutoff value was given (producing 2 stenosis ranges) for determining procedural use. 4 guidelines used different ranges of stenosis severity according to different recommended imaging techniques or procedures or the same treatment recommendations. In 2 guidelines, there were no recommendations for ACS of 50% to 60% or 69%, but there were recommendations for higher and lower degrees of ACS. Where the method of measuring ACS was indicated, it was by the NASCET method in all cases. Of 25 guidelines with CEA recommendations for patients with moderate or severe ACS (≈50%–99% by NASCET criteria), 24 (96%) endorsed CEA for average-CEA-risk patients by either recommending that it should be provided (7 guidelines) or that it may be provided (17 guidelines). In 6 guidelines, CEA endorsement for average-CEA-risk ACS was limited to patient subgroups: men with >80% stenosis, life expectancy >3 to 5 years, men <75 years, younger fitter women, high-medical-risk patients (not defined), high medical-risk because of progression of ACS, embolic signals on transcranial Doppler, history of contralateral TIAs, or silent ipsilateral cerebral infarction. Most guidelines indicated that CEA or CAS were not recommended for mild ACS (<50%–70% by NASCET) or SCS (<50% by NASCET) by not including procedural recommendations or explicitly stating that these procedures should not be done or that medical treatment alone was indicated.
What harms were identified?	A potential harm could be that all the guidelines in this review with endorsements of CEA and CAS are based on trials of CEA versus medical treatment alone, with randomized patient data from 12 to 34 years ago. There

Systematic Review	Evidence
	was a lack of evidence on stroke risk stratification for ACS.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	This is the most up-to-date review.

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

1b. Performance Gap

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

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

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

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

There is wide variation in the use of methods for stenosis calculation, which may also lead to variation in the appropriateness of carotid intervention. Since the degree of stenosis is an important element of the decision for carotid intervention, characterization of the degree of stenosis needs to be standardized. Requiring that stenosis calculation be based on a denominator of distal internal carotid diameter or, in the case of duplex ultrasound, velocity measurements that have been correlated to angiographic stenosis calculation based on distal internal carotid diameter, makes the measure applicable to both imaging and duplex studies.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis*. (**This is required for maintenance of endorsement**. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

- 2012: Performance Rate: 16.85, # of patients included: 726555, # of physicians: 3186920, Min: 0.54, Max: 100, Interquartile range: 50
- 2013: Performance Rate: 24.91, # of patients included: 769239, # of physicians: 54732, Min: 0.63, Max: 100, Interquartile range: 31.25
- 2014: Performance Rate: 81.57, # of patients included: 772456, # of physicians: 18141, Min: 0.69, Max: 100, Interquartile range: 19.35
- 2015: Performance Rate: 75.40, # of patients included: 893220, # of physicians: 13417, Min: 0.60, Max: 100, Interquartile range: 16.7, Std Deviation: 20.84
- 2016: Performance Rate: 80.16, # of patients included: 1038759, # of physicians: 15779, Min: 2.11, Max: 100, Interquartile range: 7.83, Std Deviation: 16.93
- 2017: Performance Rate: 74.49, # of patients included: 686055, # of physicians: 11387, Min: 1.27, Max: 100, Interquartile range: 4, Std Deviation: 15.14
- 2018: Performance Rate: 74.97, # of patients included: 639413, # of physicians: 9129, Min: 0.06, Max: 100, Interquartile range: 1.11, Std Deviation: 11.93

Scores by decile: Decile 3 (0.38 - 99.27), Decile 4 (99.28 - 99.83), Decile 5 (99.84 - 99.99), Decile 10 (100).

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

There is sufficient performance data.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. ***(This is required for maintenance of endorsement****. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

CMS does not provide patient information to measure stewards when providing performance data, such as race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability, for disparities analysis. The ACR has provided articles on disparities within carotid artery imaging below.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

There is limited research on disparities within carotid imaging. It is important to evaluate imaging use disparity to understand the reasons for widely reported disparities in CEA and Stenting. The ACR has identified two articles have that highlighted some disparities within carotid imaging in minority populations.

Cheng et al. (2012) conducted a retrospective cohort study on veterans hospitalized with ischemic stroke at 127 Veteran Administration (VA) hospitals in 2007. The sample consisted of 1,534 white patients and 628 black patients. Nearly 40% of all black patients were admitted to 1 of 13 minority-serving hospitals. No racial disparity in receipt of carotid artery imaging was detected within nonminority serving hospitals. However, the predicted probability of receiving carotid artery imaging for white patients at nonminority serving hospitals (89.7%, 95% CI [87.3%, 92.1%]) was significantly higher than both white patients (78.0% [68.3%, 87.8%] and black patients (70.5% [59.3%, 81.6%]) at minority serving hospitals. Cheng et al. discuss the difficulties of applying some commonly noted explanations for disparities, such as perceived risk of imaging and clinician-patient interaction (cultural competency/shared decision-making). Since carotid imaging involved very little interaction with the patient and is typically ordered without patient

input, the impact of imaging bias was greatly mitigated. The researchers do believe that site of care should be explored as an explanation of disparities by race or ethnicity if the comparison groups are obtaining medical care from different facilities. The omission of carotid artery imaging in a patient with a new ischemic stroke represents poor quality of care because eligibility for more aggressive treatment options is not ascertained.

Martin et al. (2012) conducted a study on the variation in the receipt of diagnostic carotid imaging among elderly black and white fee-for-service Medicare beneficiaries hospitalized with a primary discharge diagnosis of ischemic stroke. Patients were randomly selected; data were obtained from medical record review by two clinical data abstraction centers using computerized abstraction tools. Patient age, sex, race and medical history were recorded. A total of 19,639 elderly ischemic stroke patients were included in the analyses; 10% (n= 1,974) were identified as black, 57% were women, and the mean age was 78.2 ± 7.3 years. Black patients were more likely to be women, to be younger, and to have a history of stroke, diabetes, and/or hypertension. White patients were more likely to have prior TIA, atrial fibrillation, heart disease and/or myocardial infarction than black patients. Overall, 69.6% of patients received at least one diagnostic carotid imaging test. Duplex ultrasounds were performed in 64.7%, MRA in 11.5%, and catheter angiography in 3.4% of patients. Black ischemic stroke patients were less likely to receive diagnostic carotid imaging than white patients, although the difference was small, and only significant after risk adjustment. There was no difference in the proportion having carotid endarterectomy after adjustment for degree of carotid artery stenosis and other clinical factors. Martin et al. note that racial differences in CEA rates have been documented using Medicare administrative claims data, as well as in other national data and statewide hospital discharge information. There is greater utilization of CEA among white as compared with black patients.

Clinical characteristics that may confound the association between black and white race and receipt of the operation, such as the degree of stenosis, were not assessed. The lack of information related to the degree of stenosis in these studies may explain the discrepancy between their results and that of the present analysis.

Rather than addressing NASCET method utilization, a critical element of the measure, the papers underscore the racial/ethnic disparities associated with diagnostic imaging. Under-treatment, an implication of underdiagnosis, may result from under-utilizing clinically indicated carotid imaging and/or standardized methods for calculating the degree of stenosis. Guidelines for screening will assist with asymptomatic high-risk populations.

Cheng EM, Keyhani S, Ofner S, et al. Lower use of carotid artery imaging at minority-serving hospitals. Neurology. 2012;79(2):138-144. doi:10.1212/WNL.0b013e31825f04c5

Martin K, Naert L, Goldstein L, et al. Comparing the Use of Diagnostic Imaging and Receipt of Carotid Endarterectomy in Elderly Black and White Stroke Patients. Journal of Stroke and Cerebrovascular Diseases, Volume 21, Issue 7. 2012. https://doi.org/10.1016/j.jstrokecerebrovasdis.2011.02.002.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

De.5. Subject/Topic Area *(check all the areas that apply):*

Neurology

De.6. Non-Condition Specific*(check all the areas that apply):*

Care Coordination, Safety

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

Children, Elderly, Populations at Risk

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

https://www.acr.org/-/media/ACR/NOINDEX/Measures/2021_Measure_195_MIPSCQM.pdf

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

This is not an eMeasure **Attachment:**

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

Attachment **:** 2021_measure_195_MIPSCQM.pdf

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

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

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

Not an instrument-based measure

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

No

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

no major changes

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

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

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

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

***IF an OUTCOME MEASURE,** describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

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.

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

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

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

***IF an OUTCOME MEASURE**, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).*

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

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

No Denominator Exclusions or Denominator Exceptions

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

None

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

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

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

No risk adjustment or risk stratification

If other:

S.12. Type of score:

Rate/proportion

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Higher score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

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.

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

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

This measure is not based on a sample or survey.

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

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

N/A

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

If other, please describe in S.18.

Claims, Registry Data

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

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

Not applicable

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Clinician : Individual

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

Inpatient/Hospital, Outpatient Services

If other:

S.22. COMPOSITE Performance Measure - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

This is not a composite measure.

2. Validity – See attached Measure Testing Submission Form

NQF_Testing_Attachment_2021.docx

2.1 For maintenance of endorsement

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

Yes

2.2 For maintenance of endorsement

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

Yes

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are

not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** 0507
Measure Title: Diagnostic Imaging: Stenosis Measurement in Carotid Imaging Reports
Date of Submission: **1/1/2021**

Type of Measure:

Measure	Measure (continued)
☐ Outcome (*including PRO-PM*)	☐ Composite – *STOP – use composite testing form*
☐ Intermediate Clinical Outcome	☐ Cost/resource
☒ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	*

*cell intentionally left blank

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	Measure Tested with Data From:
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☒ registry	☒ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

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

The American College of Radiology (ACR) completed measure testing using Medicare Part B claims, qualified registry data, and qualified clinical data registry data. The data was obtained from the Centers for Medicare & Medicaid Services (CMS).

1.3. What are the dates of the data used in testing?

- The data collection period was from 2010
- The data collection period was 2012-2014

The most recent measure testing data is from January 1, 2015 - December 31, 2018

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

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

Initial Testing Project (2010):

- Three radiology practice sites representing various types, locations and sizes were identified to participate in testing the measures
- The number of physicians per site was between 10 and 1,000 physicians
- Two of the sites were hospital-based radiology group practices and one was a stand-alone radiology group practice
- All three sites were located in urban regions
- Patient visit volume ranged from 550-1600 patients, per site, per day
- Sample size included a total of 109 records for this measure
- The data collection period was 1/1/2010- 12/31/2010
- Data abstraction was performed in 2011

(2015)

The numbers of physicians were 133,717 physicians

Among these physicians 128, 525 data was based on claims and 5192 was based on registry data

(2020)

The testing sample comprised all NPIs who submitted data to CMS for this measure. The sample consisted of 55,761 physicians. The eligible population for this measure (i.e. the denominator) includes all final reports for carotid imaging studies (neck MR angiography [MRA], neck CT angiography [CTA], neck duplex ultrasound, carotid angiogram) performed. There are no exclusions to this measure.

Table 1. Number of providers that submitted data for this measure.

All- Claims, QCDR, and Registry	# of NPIs
All 4 Years	55,761
2015	15,095
2016	17,722
2017	12,713
2018	10,231
Claims	**# of NPIs**
All 4 Years	47,893
2015	12,729
2016	14,201
2017	11,759
2018	9,204
QCDR	**# of NPIs**
Both Years *	365
2015	215
2016	150
Registry	**# of NPIs**
All 4 Years	7,503
2015	2,151
2016	3,371
2017	954
2018	1,027

**CMS combined QCDR and Registry data beginning in 2017.*

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

(2015)

- Number of patients eligible were 6,877,159 (avg. per NPI is 51.43)
- Number of patients reported were 2,268,250 (avg. per NPI is 16.96)

(2020)

A total of 7,267,917 individuals were eligible to be included in this testing. However, for the Merit-based Incentives Payment System (MIPS), physicians are not required to submit all patient data to CMS. Between 2016 and 2018, a minimum 60% of data was required for reporting. The ACR performed testing with the 6,462,722 individuals that were reported to CMS.

Table 2. Eligible Patients and Reported Patients

All- Claims, QCDR, and Registry	# of Patients Eligible	# of Patients Reported
All 4 Years	7,267,917	6,462,722
2015	1,123,433	982,806
2016	1,550,485	1,387,545
2017	1,887,183	1,627,953
2018	2,706,816	2,464,418
Claims	**# of Patients Eligible**	**# of Patients Reported**
All 4 Years	3214500	2,577,610
2015	860,528	724,920
2016	949,197	802,753
2017	740,179	564,876
2018	664,596	485,061
QCDR	**# of Patients Eligible**	**# of Patients Reported**
Both Years*	269,815	257,627
2015	26,114	25,225
2016	243,701	232,402
Registry	**# of Patients Eligible**	**# of Patients Reported**
All 4 Years	3,783,602	3,627,485
2015	236,791	232,661
2016	357,587	352,390
2017	1,147,004	1,063,077
2018	2,042,220	1,979,357

** CMS combined QCDR and registry data beginning in 2017.*

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

There are no differences in the data or sample used for different aspects of testing.

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

No social risk factors were analyzed for this measure.

2a2. RELIABILITY TESTING

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

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

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

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

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

ACR performed a signal-to-noise ratio (SNR) analysis test on the performance data for reliability. In SNR analysis, reliability is the measure of confidence in differentiating performance between physicians or other providers. The signal is the variability in measured performance that can be explained by real differences in physician performance and the noise is the total variability in measured performance.

A reliability score equal to zero implies that all the variability in a measure is attributable to measurement error. A reliability score equal to one implies that all the variability is attributable to real differences in physician performance. A reliability score of 0.70 is generally considered the minimum threshold for reliability and 0.80 is generally considered very good reliability.

SNR reliability testing is performed using the Beta-Binomial Model, which assumes that physicians' performance scores are a binomial random variable conditional on the physicians' true value derived from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta are considered intermediate calculations used to establish the variance estimates.

ACR testing protocol followed the convention of estimating reliability at two points:

1) at a minimum number of qualities reporting events per physician and

2) at the average number of quality reporting events per physician. The minimum threshold of events was set at 10. Limiting the reliability analysis to physicians with a minimum number of events reduces bias introduced by the inclusion of physicians without a significant number of events.

CMS physician-level claims, registry, and QCDR data was extracted for the relevant physician-level information.

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

(2015)

The following are the results of inter-rater reliability testing.

Reliability (N, % Agreement, Kappa (95% Confidence Interval))
Overall Reliability (109, 100%, n/a*)
Numerator Reliability (109, 100%, n/a*)
Denominator Reliability (109, 100%, n/a*)

* Kappa statistics cannot be calculated because of complete agreement. Confidence intervals cannot be calculated because to do so would involve dividing by zero which cannot be done.

Summary of PQRS Reliability Score Stats Cumulative and by Year (2012 - 2014)

Year	*Number of Providers*	*Reliability p25*	*Reliability median*	*Reliability p75*	*Reliability mean*	*Reliability LCLM*	*Reliability UCLM*
2012	37142	0.81728	1	1	0.84524	0.84238	0.84809
2013	33493	0.63205	1	1	0.81781	0.81477	0.82085
2014	12953	0.99306	0.99722	0.99913	0.99473	0.99461	0.99484
All	83588	0.88581	1	1	0.85741	0.85561	0.85922

The mean (CI), P25, median, P75 of the reliability score results are shown in the above table for all 3 years as well as by each year. Our mean (CI) reliability is 0.85741 (0.85561, 0.85922). A reliability of 0.80 is considered very good reliability. So according to the reliability testing analysis, the results demonstrated very good reliability.

(2020)

Using the parameter estimates from the beta-binomial model, we computed reliability scores for each performance year. Please see **Table 3** for the results.

Table 3. Reliability Score Statistics by Year by Provider (claims and registry)

Year	Number of Providers	25th percentile	Reliability median	75th percentile	Reliability mean	Lower Confidence Limit (minimum)	Upper Confidence Limit (maximum)
2015	15095	.99467	.99804	.99950	.99593	.99584	.99602
2016	17994	.99576	.99857	.99983	.99680	.99674	.99687
2017	13579	.98857	.99601	.99957	.99185	.99167	.99203
2018	11133	.98352	.99481	.99960	.98638	.98599	.98677
ALL	57801	.99247	.99765	.99967	.99340	.99331	.99350

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

The mean (CI), P25, median, P75 of the reliability score results are shown in the above table for all 3 years as well as by each year. Our mean (CI) reliability is 0.99308 (0.99303, 0.99313). A reliability of 0.80 is considered very good reliability. Therefore, according to the reliability testing analysis, the results demonstrate very good reliability.

2020 Update:

This measure remains consistently reliable. The mean (CI) reliability is 0.99340 (0.99331, 0.99350), which is higher than the required 0.80.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Our expert panel included 14 members whose specialties include neuroradiology, abdominal radiology, musculoskeletal radiology, cardiac/thoracic radiology, breast imaging, general diagnostic radiology, nuclear medicine, informatics, quality, and physics.

- David Seidenwurm, MD, FACR (Chair) (Radiology/Neuroradiology) Sacramento, CA
- Dorothy Bulas, MD, FACR (Radiology/Pediatric Radiology) Washington, DC
- Robert Henkin, MD, FACR (Nuclear Medicine)
- Charles Johnson, MD, FACR (Radiology/Abdominal Radiology) Scottsdale, AZ
- David Rubin, MD (Radiology, Musculoskeletal Radiology) Saint Louis, MO
- Frank Rybicki, MD (Radiology/Cardiac/thoracic Radiology) Boston, MA
- Elizabeth Burnside, MD, MPH (Radiology/Breast Imaging) Madison, WI
- Matt Hawkins (Radiology Fellow) Cincinnati, OH
- Jonathan Kruskal, MBChB, PhD (Radiology/Abdominal Radiology) Newton, MA
- Frank Lexa, MD, MBA (Radiology/Neuroradiology) Wynnewood, PA
- Paul Nagy, PhD (Informatics, Quality, Physicist) Baltimore, MD
- Donald Renfrew, MD (General Diagnostic Radiology) Sturgeon Bay, WI
- Bob Pyatt, MD (General Diagnostic Radiology) Chambersburg, PA
- Paul Larson, MD (General Diagnostic Radiology) Madison, WI

This performance measure was assessed for content validity by a panel of expert work group members during the development process. Additional input on the content validity of draft measures is obtained through a 30-day public comment period and by also soliciting comments from a panel of consumer, purchaser, and patient representatives specifically for this purpose. All comments received are reviewed by the expert work group and the measures adjusted as needed. Other external review groups (e.g., focus groups) may be convened if there are any remaining concerns related to the content validity of the measures.

An expert panel was used to assess face validity of the measure. This panel consisted of 14 members, with representation from the following specialties: neuroradiology, abdominal radiology, musculoskeletal radiology, cardiac/thoracic radiology, breast imaging, general diagnostic radiology, nuclear medicine, informatics, quality, and physics.

The aforementioned panel was asked to rate their agreement with the following statement:

The scores obtained from the measure as specified will accurately differentiate quality across providers.

Scale 1-5, where 1=Strongly Disagree; 3=Neither Disagree nor Agree; 5=Strongly Agree

2020 Update:

ACR endeavored to perform construct validity empirical testing on NQF # 0507. This measure requires radiologists use a standardized, validated method for quantifying stenosis in carotid imaging studies, specifically direct or indirect measurements of the distal internal carotid diameter. The measure purpose is to improve reporting the method used to assess the degree of stenosis, Since degree is a critical factor in determining the treatment and management approach for patients with carotid stenosis, how the degree of stenosis is assessed should be standardized. By comparing NQF # 0507 performance data to data of a related measure, we intended to hypothesize that the performance of the related measure correlated with the performance of NQF #507. However, we were unable to identify a measure suitable for comparison within the

same accountability program (MIPS) for which we could obtain *patient* level data. We used the CMS Measure Repository to search for related measures.

ACR identified two measures in the MIPS program that are related to #0507: MIPS #413, *Door to Puncture Time for Endovascular Stroke Treatment*, MIPS #409, *Clinical Outcome Post Endovascular Stroke Treatment*. We also identified a related measure in the CMS Hospital Outpatient Quality Reporting (HOQR) program, NQF #0661, OP-23: *Head CT scan results for acute ischemic stroke or hemorrhagic stroke who received head CT scan interpretation within 45 minutes of arrival*. Each of these measures focus on timely and effective care for stroke care and treatment. For these measures, we were only able to obtain *population*-level measure data. NQF indicated that an acceptable alternate option for demonstrating empirical validity is to perform criterion validity using measure performance data at the *population* level. Our plan to perform the requisite analyses among these measures to determine if a relationship exists to support empirical validity, hypothesizing that hospitals or physicians performing well on these measures (MIPS #409, MIPS #413 and HOQR OP-23/NQF #0661) would perform the same on the stenosis measure (NQF #0507). However, we were unable to format the measures' data sets to perform empirical analysis. While MIPS #409 and #413 are specified at the individual clinician level, CMS was unable to provide us with individual level data, because all submissions were done at the group level. In addition, there were 26 eligible submissions between 2016 and 2018 that had performance rates. The specialties that submitted data for these measures were non-radiologists. Given the shortage of data, the group level data submission, and the different clinician types, the performance on NQF #0507 could not be compared with MIPS #409 and MIPS #413. The data from HOQR OP-23/NQF #0061 was missing the number of clinicians that performed the measure, the patient sample sizes, the numerator and denominator, so it was also unable to be correlated with NQF #0507.

Due to the lack of appropriate measurement data to perform empirical validity testing, ACR performed a new face validity survey on this measure in November 2020. An expert panel assessed the face validity of the measure. The panel consisted of 28 members in a cross-section of practice types and geographical locations.

- Amy Kotsenas, MD (Neuroradiology) Rochester, MN
- Rajeev Shah, MD, MBA (Diagnostic Radiology/Neuroradiology) Austin, TX
- Cathrine Keller, MD (Diagnostic Radiology) Leesburg, FL
- Brian Berger, MD (Diagnostic Radiology) Shelby Township, MI
- Yvonne Lui, MD (Neuroradiology/AI) New York, NY
- Ajay Gupta, MD (Diagnostic Radiology/Neuroradiology) New York, NY
- Haris Sair, MD (Diagnostic Radiology/AI) Baltimore, MD
- Jeffrey Jarvik, MD, MPH (Neuroradiology) Seattle, WA
- Max Wintermark, MD (Neuroradiology) Stanford, CA
- Michael Iv, MD (Neuroradiology) Palo Alto, CA
- Jeffrey Stone, MD (Diagnostic Radiology) Jacksonville, FL
- Theodore Larson III, MD (Interventional Radiology/Neuroradiology) Centennial, CO
- Steven Falcone, MD (Neuroradiology) Miami, FL
- Gloria Guzman, MD, MSc, MPH (Neuroradiology) St. Louis, MO
- Jody Tanabe, MD (Diagnostic Radiology/Neuroradiology) Aurora, CO
- Achala Vagal, MD (Neuroradiology) Cincinnati, OH
- Sammy Chu, MD (Diagnostic Radiology) Bellevue, WA
- Nolan Kagetsu, MD (Diagnostic Radiology) New York, NY
- Bradley Delman, MD (Neuroradiology) New York, NY

- John Jordan, MD (Neuroradiology) Torrance, CA
- Fabio Settecase, MD (Diagnostic Radiology) San Francisco, CA
- Patrick Turski, MD (Diagnostic Radiology/Medical Physics) Madison, WI
- Mariya Gusman (Neuroradiology Fellow) Fairfield, CA
- Jacob Ormsby, MD, MBA (Diagnostic Radiology/Neuroradiology) Albuquerque, NM
- William Donovan, MD, MPH (Neuroradiology) Norwich, CT
- Noushin Yahyavi Firouz Abadi, MD (Neuroradiology) Potomac, MD
- Roland Lee, MD (Neuroradiology) San Diego, CA
- Salil Soman, MD (Diagnostic Radiology/Neuroradiology) Boston MA

The panel was asked to rate their agreement with the following statement: The scores obtained from the measure, as specified, will accurately differentiate quality across providers. The panel could choose from a scale of 1-5, where 1=Strongly Disagree; 3=Neither Disagree nor Agree; 5=Strongly Agree.

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

The results of the expert panel rating of the validity statement were as follows: N = 7; Mean rating = 4.43 and 85.71% of respondents either agree or strongly agree that this measure can accurately distinguish good and poor quality

This measure underwent maintenance review by an expert panel. The review was completed in February 2015. New evidence was reviewed. The expert panel agreed that the measure remained valid based on existing and new evidence.

82.15% (23 members) of the panel either strongly agreed or agreed that this measure accurately distinguishes good from poor quality. Two panel members disagreed that the measure would accurately distinguish good from poor quality. One member stated that literature shows that CTA underestimates the stenosis and MRA overestimates the stenosis, compared to NASCET. The other member stated that ultrasound should be removed from the measure. He also added that stenosis on ultrasound is measured using velocities and NASCET should never be applied to ultrasound. However, the ACR respectfully disagrees and believes perhaps the commenter may not have fully understood the goal or construct of the measure. The types of imaging included in the measure (MRA, CTA, duplex US, angiography) are all used in practice to evaluate carotid artery stenosis. The intent of the measure is to improve consistency in reporting the method used to estimate stenosis, agnostic to the modality/technology of the carotid imaging study used to evaluate level of stenosis. Evidence shows that NASCET has standardized the method of quantifying stenosis, specifically with reference to the distal internal carotid diameter as the denominator for stenosis measurement, in several different types of imaging studies. There are validated methods to cross reference and correlate indirect reference to the carotid diameter to angiography or other imaging studies, which are gold standard. For example, in Doppler ultrasound, the degree of stenosis can be estimated using Doppler parameter of the peak systolic velocity (PSV) of the internal carotid artery (ICA), with concordance of the degree of narrowing of the ICA lumen. Additional Doppler parameters of ICA-to-common carotid artery (CCA) PSV ratio and ICA end-diastolic velocity (EDV) can be used when degree of stenosis is uncertain from ICA PSV. (Grant et al, Society of Radiologists in Ultrasound, 2003).

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)
This measure remains valid.

_____

2b2. EXCLUSIONS ANALYSIS

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

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis.* ***Note**: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

_____

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

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

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

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

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

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

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

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

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

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

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

2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b3.11. Optional Additional Testing for Risk Adjustment (***not required****, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

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

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To assess statistically significant differences in measure rates, the data described in sections above were used to calculate the mean, median, standard deviation, and interquartile range for the measure rates. In addition, the rates were divided into quartiles, and a Student's t-test was used to compare the rates of the plans in the 25th percentile to the rates of the plans in the 75th percentile.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

The tables below show the distribution of measure rates for **claims** data between 2015 and 2018. The mean rate was 94.52%, with a median rate of 100%, minimum rate of 1.27%, and maximum rate of 100%.

Table 4. Variation in Measure Rates for Claims Data – 2015 to 2018

Mean	Median	Standard Deviation
94.52%	100%	14.17%

Table 5. Distribution of Measure Rates for Claims Data – 2015 to 2018

Statistic	Value
Minimum	1.27%
25th percentile	97.22%
50th percentile (median)	100%
75th percentile	100%
Maximum	100%
Interquartile Range	2.78%
Student's t-test p-value	P<0.0001

The tables below show the distribution of measure rates for **Registry** data between 2015 and 2018. The mean rate was 98.48%, with a median rate of 100%, minimum rate of 0.06%, and maximum rate of 100%.

Table 6. Variation in Measure Rates for Registry Data – 2015 to 2018

Mean	Median	Standard Deviation
98.48%	100%	6.94%

Table 7. Distribution of Measure Rates for Registry Data – 2015 to 2018

Statistic	Value
Minimum	0.06%
25th percentile	100%
50th percentile (median)	100%
75th percentile	100%
Maximum	100%
Interquartile Range	0 %
Student's t-test p-value	P<0.0001

The tables below show the distribution of measure rates for **QCDR** data between 2015 and 2016. As a reminder, the QCDR data for 2017 and 2018 is combined in the registry data above. The mean rate was 293.45%, with a median rate of 100 %, minimum rate of 0%, and maximum rate of 100%.

Table 8. Variation in Measure Rates for QCDR Data – 2015 to 2016

Mean	Median	Standard Deviation
97.62%	100%	8.35%

Table 9. Distribution of Measure Rates for QCDR Data – 2015 to 2016

Statistic	Value
Minimum	18.18%
25th percentile	100%
50th percentile (median)	100%
75th percentile	100%
Maximum	100%
Interquartile Range	0%

Statistic	Value
Student's t-test p-value	P<0.0001

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

For all sets of data, the measure rates did not show significant variation in the interquartile ranges, but did show a statistically significant difference in the measure rates between the top and bottom quartile of the plans included in the testing (P<0.0001 at alpha = 0.05).

Submission Method	Interquartile Range
Claims	2.78%
Registry	0%
QCDR	0%

However, while the variation in the data set appears low, the ACR reviewed the number of eligible instances where physicians could have submitted this measure against the number of instances reported to CMS. The performance rate is high among physicians who chose to report this measure to CMS. However high-performing measures may have low adoption rates among all physicians. High performance scores may result from a small pool of high-performing physicians who report such measures, thereby potentially underestimating the extent of variation of the measure action across physicians. Based on the large discrepancy in the reporting rate for this measure when comparing CMS claims and registry data submissions to the number of eligible reporting instances (based on billed CPT codes for exams relevant to the measure denominator), this measure may have a larger performance gap than the CMS reporting rate shows.

Table 10. Reporting rate for individuals that submitted to CMS vs. all reporters

Measures	Reporting rate across all individuals who reported the measure to CMS	Reporting Rate for all eligible reporting instances
2015	88%	35%
2016	90%	48%
2017	86%	55%
2018	91%	84%

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

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

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

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

With the use of claims and registry as the data sources for this measure, CMS Medicare and Medicaid administrative data is valid and reliable since it determines eligibility for enrollment and payment of services. Registry data submissions may have some missing data, as registry users are not required to submit *all* data to CMS. Registry users are required to submit 70% of their data. However, the volume of patients (6,462,722) used in this data set greatly minimizes the risk of bias.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse;* ***if no empirical sensitivity analysis****, identify the approaches for handling missing data that were considered and pros and cons of each*)

Missing data related to registry data providers not submitting information on patients was previously noted. However, the number of patients that were eligible (7,267,917) compared to the amount submitted and used for this analysis (6,462,722) likely would not have made a significant difference in the testing results. It does, however, make a difference in the performance gap for this measure. The performance may be affected by the lack of responses.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted;* ***if no empirical analysis****, provide rationale for the selected approach for missing data*)

The performance results are from a significantly large data set of over 6,000,000 patients. The loss of about 800,000 eligible patients would be unlikely to create a bias or a significant difference in the results. Yearly, CMS raises the volume of data required for submission in the MIPS program. This will assist with minimizing bias even more in the future.

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

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

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

Some data elements are in defined fields in electronic sources

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

The data elements are manually abstracted from the radiology report. The ACR is working to enable extraction of free text from radiology reports using Artificial Intelligence (AI) and Natural Language Processing (NLP).

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For

eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

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

During the measures' operational use, some users reported minor difficulties regarding the numerator, (i.e. data needed to meet the numerator). Updates clarifying the measure's definitions and instructions for capturing the numerator were incorporated, based on feedback from the MIPS program.

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

Not applicable.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting	Payment Program Merit Based Incentives Payment System qpp.cms.gov Merit Based Incentives Payment System qpp.cms.gov Quality Improvement (Internal to the specific organization) ACR Quality Improvement nrdr.acr.org

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

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

This measure is used in the CMS Payment Program (Merit-based Incentive Payment System) for accountability and reimbursement. Over 10,000 physicians and approximately 2.4 million patients are included in the

program for this measure. A variety of geographic areas in the United States are measured. Measurement is performed at the individual and group levels.
This measure is also used for quality improvement within the ACR registries.

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
This is an accountability measure and used in the CMS quality and payment programs.

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

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

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

The measure specifications are updated annually and are included in the CMS Quality Payment Program for MIPS. The measure is reported via claims and registry as MIPS # 195 or Quality ID #195. Detailed specifications are publicly available on the CMS resource library.

Assistance with interpretation for this measure is provided through the ACR help desk and through the CMS help desk. Users can submit their questions and receive a response from ACR staff within 72 hours

Performance results are provided in two ways. First, through Qualified Clinical Data Registries (QCDRs). Users upload the measure data to the OCDR. Quarterly, measure users may compare their performance on this measure against CMS performance benchmarks. To view performance results online, users must have an active account within the QCDR. The second method for which data is provided is through CMS' annual MIPS Feedback Reports. The feedback reports, aggregated at a high-level, are also based on CMS performance benchmarks (calculated in deciles). CMS Feedback Reports are nonspecific and not necessarily indicative of an individual clinician's performance.

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

Feedback is provided quarterly to all QCDR participants reporting this quality measure. Feedback is based on CMS performance benchmarks, which are calculated in deciles. These reports are nonspecific and not necessarily indicative of an individual clinician's performance.

ACR educational webinars are conducted bimonthly to explain measure requirements and interpretation of performance results.

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

Describe how feedback was obtained.

Feedback is obtained through email, the ACR help desk, the CMS quality help desk, and CMS contractor QMMS. Feedback has been positive.

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

Feedback on this measure has been primarily clarifying questions on how to report certain cases, such as ones with no stenosis. Overall, radiologists agree that having a standardized method for calculating stenosis is a valuable tool in stroke imaging.

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

No other feedback has been provided from entities other than individuals that could report the measure.

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

This feedback is considered during the annual measure specification update process with CMS. The ACR Metrics Committee reviews feedback for measure changes.

Improvement

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

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

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

Performance data shows significant improvement for this measure.

4b2. Unintended Consequences

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

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

We are not aware of any unintended consequences related to this measurement.

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

Implementing this measure has created more standardization for carotid imaging results. This measure also supports communication between radiologists and referring physicians.

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

No

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

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

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?
5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.
5b. Competing Measures

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

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Appendix

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

Available at measure-specific web page URL identified in S.1 **Attachment:**

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American College of Radiology (ACR)

Co.2 Point of Contact: Karen, Campos, kcampos@acr.org, 800-227-5463-5848

Co.3 Measure Developer if different from Measure Steward: American College of Radiology (ACR)

Co.4 Point of Contact: Karen, Campos, kcampos@acr.org, 800-227-5463-5848

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

List of Work Group Members:

William Golden, MD (Co-Chair) (internal medicine)

David Seidenwurm (Co-chair) (diagnostic radiology)

Michael Bettmann, MD

Dorothy Bulas, MD (pediatric radiology)

Rubin I. Cohen, MD, FACP, FCCP, FCCM

Richard T. Griffey, MD, MPH (emergency medicine)

Eric J. Hohenwalter, MD (vascular interventional radiology)

Deborah Levine, MD, FACR (radiology/ultrasound)

Mark Morasch, MD (vascular surgery)

Paul Nagy, MD, PhD (radiology)

Mark R. Needham, MD, MBA (family medicine)

Hoang D. Nguyen (diagnostic radiology/payer representative)

Charles J. Prestigiacomo, MD, FACS (neurosurgery)

William G. Preston, MD, FAAN (neurology)

Robert Pyatt, Jr., MD (diagnostic radiology)

Robert Rosenberg, MD (diagnostic radiology)

David A. Rubin, MD (diagnostic radiology)

B Winfred (B.W.) Ruffner, MD, FACP (medical oncology)

Frank Rybicki, MD, PhD, FAHA (diagnostic radiology)

Cheryl A. Sadow, MD (radiology)

John Schneider, MD, PhD (internal medicine)

Gary Schultz, DC, DACR (chiropractic)

Paul R. Sierzenski, MD, RDMS (emergency medicine)

Michael Wasylik, MD (orthopedic surgery)

Diagnostic Imaging Measure Development Work Group Staff

American College of Radiology: Judy Burleson, MHSA; Alicia Blakey, MS

American Medical Association-convened Physician Consortium for Performance Improvement: Mark Antman, DDS, MBA; Kathleen Blake, MD, MPH; Kendra Hanley, MS; Toni Kaye, MPH; Marjorie Rallins, DPM; Kimberly Smuk, RHIA; Samantha Tierney, MPH; Stavros Tsipas, MA

National Committee for Quality Assurance: Mary Barton, MD

PCPI measures are developed through cross-specialty, multi-disciplinary work groups. All medical specialties and other health care professional disciplines participating in patient care for the clinical condition or topic under study must be equal contributors to the measure development process. In addition, the PCPI strives to include on its work groups individuals representing the perspectives of patients, consumers, private health plans, and employers. This broad-based approach to measure development ensures buy-in on the measures from all stakeholders and minimizes bias toward any individual specialty or stakeholder group. All work groups have at least two co-chairs who have relevant clinical and/or measure development expertise and who are responsible for ensuring that consensus is achieved and that all perspectives are voiced.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2007

Ad.3 Month and Year of most recent revision: 08, 2020

Ad.4 What is your frequency for review/update of this measure? These measures are updated each year.

Ad.5 When is the next scheduled review/update for this measure? 08, 2021

Ad.6 Copyright statement: COPYRIGHT:

The Measure is not a clinical guideline, does not establish a standard of medical care, and has not been tested for all potential applications.

The Measure, while copyrighted, can be reproduced and distributed, without modification, for noncommercial purposes, e.g., use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of the Measure for commercial gain, or incorporation of the Measure into a product or service that is sold, licensed or distributed for commercial gain.

Commercial uses of the Measure require a license agreement between the user and the American College of Radiology (ACR). Neither ACR nor its members shall be responsible for any use of the Measure.

The PCPI's and AMA's significant past efforts and contributions to the development and updating of the Measures are acknowledged. ACR is solely responsible for the review and enhancement ("Maintenance") of the Measure as of August 1, 2020.

ACR encourages use of the Measure by other health care professionals, where appropriate.

THE MEASURE AND SPECIFICATIONS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND.

Limited proprietary coding may be contained in the Measure specifications for convenience. A license agreement must be entered prior to a third party's use of Current Procedural Terminology (CPT ®) or other proprietary code set contained in the Measures. Any other use of CPT or other coding by the third party is strictly prohibited. ACR and its members disclaim all liability for use or accuracy of any CPT or other coding contained in the specifications.

Ad.7 Disclaimers: See copyright statement above.

Ad.8 Additional Information/Comments: Coding/Specifications updates occur annually. The ACR has a formal measurement review process that stipulates regular (usually on a three-year cycle, when feasible) review of the measures. The process can also be activated if there is a major change in scientific evidence, results from testing or other issues are noted that materially affect the integrity of the measure. Additionally, this measure is updated annually for coding changes and reviewed by CMS´ contractor QMMS.