NATIONAL QUALITY FORUM
Driving measurable health improvements together

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

NQF #: 2797
Measure Title: Transcranial Doppler Ultrasonography Screening Among Children with Sickle Cell Anemia
Measure Steward: Q-METRIC – The University of Michigan
sp.02. Brief Description of Measure: The percentage of children ages 2 through 15 years old with sickle cell anemia (Hemoglobin SS) who received at least one transcranial Doppler (TCD) screening within a year.
1b.01. Developer Rationale: Children with sickle cell anemia (HbSS) have over three hundred times the stroke risk than children with normal hemoglobin (Verduzco and Nathan, 2009). Without intervention, approximately 11% of children with sickle cell anemia will have a stroke by age 20 (Verduzco and Nathan, 2009; Ohene-Frempong et al., 1998). Transcranial Doppler (TCD) ultrasonography measures the blood velocities within the cerebral vessels (Adams et al., 1997; Adams et al., 1992). Children over the age of 2 with a time-average mean maximum blood flow velocity of 200cm/sec or greater as measured by TCD ultrasonography have been shown to have 27 times the risk of stroke than children with velocities less than 200cm/sec. This corresponds to a 40% risk of stroke among those with high velocities within 3 years (Adams et al., 1997). Initiation of chronic blood transfusions reduces the risk of stroke by 92% among children at highest risk of stroke as identified through TCD screening (Adams et al., 1997; Adams et al., 1992). TCD screening is a reasonable method to assess stroke risk among children with sickle cell anemia, as it is safe, non-invasive and low cost (Markus, 2000). Although other predictors of stroke have been examined, such as hematocrit levels and white blood cell count, TCD velocities have been shown to be the only independent predictor of stroke (Adams et al., 1992). Given the importance of TCD screening to stroke prevention among children with sickle cell anemia, the National Heart, Lung, and Blood Institute (NHLBI) recommends each child with sickle cell anemia receive one TCD screen per year from ages 2 to 16 years (National Heart, Lung, and Blood Institute, 2014). Although the benefits of TCD screening among children with sickle cell anemia have been known since the late nineties, prior studies indicate that TCD screening rates are low. However, these reports are limited in their generalizability, as they are often focused on a single healthcare provider or registry. This measure establishes a claims-based method for identifying receipt of TCD screening among larger and broader populations of children with sickle cell anemia. The measure specifications are reflective of the guidelines from the NHLBI, and the performance scores calculated through this measure will identify areas in need of improvement in receipt of TCD screening among children with sickle cell anemia.
Citations:
Adams RJ, McKie VC, Carl EM, et al. Long-term stroke risk in children with sickle cell disease screened with transcranial Doppler. Ann Neurol. Nov 1997;42(5):699-704.
Adams R, McKie V, Nichols F, et al. The use of transcranial ultrasonography to predict stroke in sickle cell disease. N Engl J Med. Feb 27 1992;326(9):605-610.
Markus HS. Transcranial Doppler ultrasound. Br Med Bull. 2000;56(2):378-388.

National Heart Lung and Blood Institute. Evidence Based Management of Sickle Cell Disease. 2014; http://www.nhlbi.nih.gov/health-pro/guidelines/sickle-cell-disease-guidelines/sickle-cell-disease-report.pdf. Accessed 11/11, 2014.
Ohene-Frempong K, Weiner SJ, Sleeper LA, et al. Cerebrovascular accidents in sickle cell disease: rates and risk factors. Blood. Jan 1 1998;91(1):288-294.
Verduzco LA, Nathan DG. Sickle cell disease and stroke. Blood. Dec 10 2009;114(25):5117-5125.

---

sp.12. Numerator Statement: The numerator is the number of children ages 2 through 15 years old with sickle cell anemia who received at least one TCD screening within the measurement year.
sp.14. Denominator Statement: The denominator is the number of children ages 2 through 15 years with sickle cell anemia within the measurement year.
sp.16. Denominator Exclusions: There are no denominator exclusions.

Measure Type: Process
sp.29. Data Source: Claims
sp.07. Level of Analysis: Health Plan

IF Endorsement Maintenance – Original Endorsement Date: 5/04/2016
Most Recent Endorsement Date: 5/04/2016

IF this measure is included in a composite, NQF Composite#/title:
IF this measure is paired/grouped, NQF#/title:
sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:

Preliminary Analysis: Maintenance of Endorsement

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

Criteria 1: Importance to Measure and Report

1a. Evidence

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

1a. Evidence. The evidence requirements for a ***structure, process or intermediate outcome*** measure are 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 description for this measure:

- This is a maintenance process measure at the health plan level that assesses the percentage of children ages 2 through 15 years old with sickle cell anemia (Hemoglobin SS) who received at least one Transcranial Doppler (TCD) screening within a year.

- The developer provides a logic model that depicts the importance of TCD screening and primary stroke prevention among children between the ages of 2-15 years old diagnosed with Sickle Cell Anemia (SCA).

The developer provides the following evidence for this measure:

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

Summary of prior review in 2016

- The developer presented clinical practice guidelines from the 2014 National Heart, Lung, and Blood Institute (NHLBI) that support annual TCD screening among children diagnosed with SCA between the ages of 2 and 15 years old
    - The developer reported a total of 2 randomized control trials (RCTs) and 50 observational studies are included in the body of evidence.
    - The developer noted that the evidence to support annual TCD screenings in children diagnosed with SCA is graded as Strong Recommendation with Moderate Quality of Evidence.
- The Standing Committee noted the measure's age specification slightly differed from the practice guideline but, overall, agreed that the measure aligns with the NHLBI guidelines for annual transcranial doppler (TCD) screening of children with sickle cell anemia.

Changes to evidence from last review

☒ The developer attests that there have been no changes in the evidence since the measure was last evaluated.

☐ The developer provided updated evidence for this measure:

Exception to evidence

- NA

Questions for the Committee:

- *The developer attests the underlying evidence for the measure has not changed since the last NQF endorsement review. Does the Committee agree the evidence basis for the measure has not changed and there is no need for repeat discussion and vote on Evidence?*

Guidance from the Evidence Algorithm

Not a health outcome or PRO (Box 1) -> Process measure based on systematic review (Box 3) -> QQC presented (Box 4) ->Quantity: High; Quality: Moderate; Consistency: Moderate/High (Box 5) -> Moderate (Box 5b) -> Moderate

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

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

Maintenance measures – increased emphasis on gap and variation

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

- The developer provided individual and overall TCD screening rates for six states from 2005 to 2010 for children diagnosed with SCA.
    - Louisiana: 41 percent to 51 percent
    - Texas 5 percent to 40 percent
    - Florida 21 percent to 46 percent
    - Illinois 25 percent to 31 percent
    - Michigan 11 percent to 43 percent
    - South Carolina 19 percent to 51 percent
    - Total rates: 22 percent to 32 percent
- The developer also provided TCD screening rates for children enrolled in two state Medicaid programs from 2011 through 2018.
    - TCD screening rates in Michigan ranged from 35.55 percent to 47.16 percent with an overall rate of 41.58 percent.
    - TCD screening rates in New York State ranged from 39 percent to 45 percent with an overall rate of 42 percent.

Disparities

- The developer noted evaluated whether there were disparities in care based on gender, insurance, socioeconomic status or age.
    - While data do not show disparities in care based on gender, insurance or socioeconomic status, the developer noted that children between the ages of 2 to 6 years old had a higher TCD screening rate (36 percent) compared to older children ages 7 to 11 years old (31 percent) and 12 to 15 year old (25 percent).

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

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

Committee Pre-evaluation Comments:

1a. Evidence

- No changes to evidence.
- The developer provided evidence of a systematic review of evidence. The clinical practice guideline from the NHLBI was provided that supported annual TCD screening among children diagnosed with SCA between the ages of 2 and 15 years old. Evidence included two randomized control trials and 50 observational studies. The evidence provided was graded as Strong Recommendation with Moderate Quality of Evidence. The measure's age specification varies slightly from the practice guideline. The developer noted that there have not been any changes in the evidence since it was last evaluated. No need to repeat the discussion on evidence.
- Moderate level evidence.

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

- Gap persists.

- The developer provided overall and individual screening rates for six states from 2005 to 2010. The screening rates ranged from a low of 5% in Texas and a high of 51% in Louisiana and South Carolina. Screening rates for two Medicaid programs. Michigan ranged from 35.55% to 47.16% with an overall rate of 41.58%. In New York the Medicaid rate ranged from 39% to 45% with an average overall rate of 42%. There were no disparities noted other than children ages 2 to 6 years having a higher TCD screening rate (36%) compared to older children 7 to 11 years old (31%) and 12 to 15 years (25%).
- Opportunity for improvement exists.

Criteria 2: Scientific Acceptability of Measure Properties

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

Evaluators: NQF Staff

2a. Reliability: Specifications and Testing

For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

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

For maintenance measures – less emphasis if no new testing data provided.

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

Specifications:

- Measure specifications are clear and precise.

Reliability Testing:

- Reliability testing conducted at the Accountable Entity Level:
    - The developer conducted empirical reliability testing at the accountable entity level using a signal-to-noise analysis to distinguish the performance of one state's Medicaid program from that of another state.
    - The developer notes data was collected from the Centers for Medicare & Medicaid Services (CMS) Analytic eXtract (MAX) administrative claims data from 2005 to 2012 for six state Medicaid programs
    - The developer presented reliability statistics ranging from 0.96 - 0.99 (median 0.98).
    - The developer indicates this is a high degree of reliability.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure cannot be consistently implemented (i.e., are measure specifications adequate)?*
- *The developer attests the specifications have not changed and that additional reliability testing was not conducted. Does the Committee agree that the measure is still reliable and there is no need for repeat discussion and vote on Reliability?*

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

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

For maintenance measures – less emphasis if no new testing data provided.

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.

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

Validity Testing

- Validity testing conducted at the Patient/Encounter Level:
    - Inter-rater reliability of paper records was conducted by examining 10 charts; the two trained abstractors had 100% agreement for receipt of TCD screening from the medical records, resulting in a kappa of 1.00
    - The developer conducted denominator validation testing using both ICD-9 and ICD-10 diagnosis codes.
    - The developer reported sensitivity and specificity values for ICD-9 claims codes as 91.4 percent and 80 percent, respectively (PPV: 80.4%; NPV: 91.3%).
    - The developer reported sensitivity and specificity values for ICD-10 claims code ranges from 0.02-0.97 and 0.88-1.0, respectively.
    - The developer noted the agreement for TCD screening rates between Michigan Medicaid claims and medical record data from 2005 to 2010 was 96.7% (kappa = 0.93, 95% confidence interval (CI): 0.86) and considered this result as perfect agreement.
- Validity testing conducted at the Accountable Entity Level:
    - The developer conducted empirical validity testing by calculating the rate of TCD screening for Michigan MAX data from 2007-2009 and the Michigan Medicaid claims data warehouse using z-tests.
        - 2007 results: z-score= -0.685, p-value= 0.497
        - 2008 results: z-score= 0.223, p-value= 0.223
        - 2009 results: z-score= 1.079, p-value=0.280
    - The developer noted that, compared to the cold standard of Michigan Medicaid data, the MAX data has a very high degree of validity. Additionally, the developer reported that the high values of the Pearson correlation coefficient (0.98) and the squared coefficient (0.96) indicate a high level of reliability.
    - Face validity was established by a panel of 14 national experts and parent advocates, as well as measurement and state Medicaid experts.
- The expert panel agreed that this measure has a high degree of face validity and rated this measure 8.5 out of 9. The panel concluded that this measure would improve the quality of care provided to patients with SCA and would be able to distinguish good from poor quality.

Exclusions

- The measure does not use exclusions.

Risk-Adjustment

- The measure is not risk adjusted or stratified.

Meaningful Differences

- The developer estimated the percentage of children with SCA that received at least one TCD screening from 2005 to 2010 using Medicaid MAX data across six state programs.
    - The developer noted that children had a higher odd of receiving TCD screening in 2007 compared to those children in 2005 (p=0.0001).
- The developer indicated that the measure can distinguish differences in performance across years and detect changes over time.

Missing Data

- Given that this measure uses administrative data, the developer did not test the extent of non-response.
- The developer states that this measure is subject to all the usual limitations of administrative data and does not anticipate data would be missing differentially across any groups.

Comparability

- The measure only uses one set of specifications for this measure.

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*

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

Committee Pre-evaluation Comments:

2a. Reliability-Specification

- This can be consistently implemented.
- The measure specifications are clear and precise.
- No concerns.

2a2. Reliability-Testing

- None.
- The developer conducted empirical reliability testing at the accountable entity level using a signal-to-noise analysis. The developer used data collected from CMS's Analytic eXtract (MAX) administrative claims data from 2005 to 2012 for six State Medicaid programs. Reliability statistics ranged from 0.96 to0.99 (median 0.98). The developer noted that this indicates a high degree of reliability.
- No concerns.

2b. Validity-Testing

- None.
- The developer conducted interrater reliability of paper records (10 charts) and found 100% agreement for receipt of TCD screening from the medical records resulting in a kappa of 1.0Sensitivity and specificity values for ICD-9 claim codes was 91.4% and 80%. ICD-10 claims code ranges from 0.02-0.97 and 0.88-1.0. TCD screening rates between Michigan Medicaid claims and medical record data was 96.7% (Kappa-0.93) with a 95% confidence interval - considered the result perfect agreement. The measure has a high degree of face validity and the expert panel rated the measure at 8.5 out of 9. The measure could distinguish good care from poor quality care.
- No concerns.

2b2-2b3. Potential threats to validity

- No concerns.
- There were no other threats to validity noted. The measure does not include risk adjustment. There are no exclusions noted in the measure specifications.
- No concerns.

2b4-2b7. Potential threats to validity

- No concerns.
- The measure uses administrative data, therefore missing data was not considered a concern for reporting. There were no concerns noted with identification of meaningful differences or the comparability of performance score. There was only one set of measure specifications.
- No concerns.

Criterion 3. Feasibility

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

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

- The developer noted that although availability of administrative claims data may vary by source (e.g., State Medicaid versus National Medicaid), all data elements are in defined fields in electronic claims.
- The developer noted that data elements needed to compute the measure score can be coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims).

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Are the required data elements available in electronic form, e.g., EHR or other electronic sources?*
- *Is the data collection strategy ready to be put into operational use?*

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

Committee Pre-evaluation Comments:

3. Feasibility

- No issues.
- All data is available in defined fields in electronic claims. The data source is administrative data/claims data. The data elements are routinely generated and used during care delivery. The data source is available and accessible. The measure is ready to be operationalized.
- Moderate feasibility.

Criterion 4: Use and Usability

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 evaluates 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
Planned use in an accountability program?	☒	Yes	☐	No	☐	NA

Accountability program details

- The measure is used within the Michigan Medicaid program (payment program) to improve rates of transcranial doppler screening among children with sickle cell anemia in southeast Michigan.
- The measure is currently not publicly reported; however, the developer indicated there are plans for the measure to be publicly reported and used within quality improvement.

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 provides consultation to the Michigan Department of Health and Human Services as this measure is integrated into their Medicaid reporting systems.
- The developer provides assistance with measure specification interpretation to those who are interested in using the measure.
- The developer provides results to Michigan Medicaid health plans on a quarterly basis.

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*
- *How has those being measured and other users been given an opportunity to provide feedback on the measure performance or implementation?*
- *How has this feedback been considered when changes are incorporated into the measure?*

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

4b. Usability (4b1. Improvement; 4b2. Benefits of measure)

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

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

Improvement results

- The developer noted that measure results are provided quarterly to Michigan Medicaid health plans; however, the developer also noted that the Medicaid health plans have recently begun quality improvement efforts and that improvement is not yet apparent.

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 did not identify any unexpected findings.

Potential harms

- No potential harms were identified by the developer.

***Questions for the Committee**:*

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

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

Unable to assess usability as there was no performance data submitted for this measure.

Committee Pre-evaluation Comments:

4a. Use

- Not publicly reported.
- The measure is not currently publicly reported. The measure is used within the Michigan Medicaid program (payment program). The developer noted that there are plans for the measure to be publicly reported and used within quality improvement. The developer did not indicate in which programs. The developer provides consultation to the Michigan Department of Health and Human Services for their reporting program. Results are provided by the developer to Michigan Medicaid health plans on a quarterly basis. Use of the measure provides an opportunity to further the goal of high-quality, efficient healthcare.
- Currently in use for accountability, but not publicly reported.

4a. Usability

- No performance data submitted. No harms identified.
- There were not potential harms noted by the developer. The measure is specific to children diagnosed with sickle cell anemia (population). The Michigan health plans have begun quality improvement efforts to improve the rate.
- No concerns.

Criterion 5: Related and Competing Measures

Related measures

- NQF #3166 Antibiotic Prophylaxis Among Children with Sickle Cell Anemia
- NQF #3595 Hydroxyurea Use Among Children with Sickle Cell Anemia

Harmonization

- The developer indicates that measures are harmonized to the extent possible.

Committee Pre-evaluation Comments:

5: Related and Competing Measures

- Related measures are harmonized to extent possible.
- Related measures include NQF #3166 Antibiotic Prophylaxis Among Children with Sickle Cell Anemia; and NQF #3595 Hydroxyurea Use Among Children with Sickle Cell Anemia. The developer noted that the measure has been harmonized to the extent possible.
- No concerns.

Public and NQF Member Comments (Submitted as of June 14, 2022)

Member Expression of Support

- No members submitted an expression of support for this member.

Comments

- No NQF member and public comments were received in advance of the Standing Committee evaluation.

Scientific Acceptability Evaluation

RELIABILITY: SPECIFICATIONS

1. **Have measure specifications changed since the last review? ☐ Yes ☒ No**
2. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented? ☒ Yes ☐ No**
3. **Briefly summarize any changes to the measure specifications and/or concerns about the measure specifications.**
   - No changes were made to the measure specifications.

RELIABILITY: TESTING

4. **Did the developer conduct new reliability testing? ☐ Yes ☒ No**

   4a. If no, summarize the Standing Committee's previous feedback:
   - The Standing Committee expressed concern about the process of identifying children with sickle cell disease and noted the measure should include stringent diagnosis specifications for identifying the condition, but ultimately agreed the developer demonstrated that patients with sickle cell disease could be reliably identified

   4b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:
   - N/A
5. **Reliability testing level: ☒ Accountable-Entity Level ☐ Patient/Encounter Level ☐ Neither**
6. **Reliability testing was conducted with the data source and level of analysis indicated for this measure:**

   ☒ Yes ☐ No

7. If accountable-entity level and/or patient/encounter level reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical VALIDITY testing** of patient-level data conducted?

   ☐ **Yes** ☐ **No**

8. **Assess the method(s) used for reliability testing:**
   - The reliability was conducted using a signal-to-noise analysis that focused on distinguishing the performance of one state's Medicaid program to another through an estimation using a beta-binomial model.
9. **Assess the results of reliability testing**
   - Reliability based on signal-to-noise analysis ranged from 0.96 to 0.99 with a median of 0.98. Values above 0.9 are considered sufficient to see differences between states resulting in a high degree of reliability.
10. 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.

    ☒ **Yes** ☐ **No** ☐ **Not applicable**

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

    ☒ **Yes** ☐ **No** ☐ **Not applicable** (patient/encounter level testing was not performed)

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

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

    ☐ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity 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)

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

    Measure specifications precise, unambiguous, and complete (Box 1) -> Empirical reliability testing conducted with the measure as specified (Box 2) -> Empirical testing at the accountable entity level (Box 4) -> Reliability testing method described and appropriate (Box 5) -> High certainty or confidence that the levels are reliable (Box 6a) -> High rating

VALIDITY: TESTING

14. **Did the developer conduct new validity testing? ☒ Yes ☐ No**

    14a. If no, summarize the Standing Committee's previous feedback:
    - N/A

    14b. If yes, describe any differences between the new and old testing and summarize any relevant Standing Committee's feedback from the previous review:
    - The addition of ICD-10-CM along with ICD-9-CM increases identification of children with sickle cell anemia. The ICD-10-CM denominator validation was developed, tested, and validated the performance through a four-step process: developing candidate SCA case definitions; identifying a test population for the definitions; testing the accuracy of candidate case definitions; and testing case definitions within an independent population.

15. **Validity testing level (check all that apply):**

    ☐ **Accountable-Entity Level** ☐ **Patient or Encounter-Level** ☒ **Both**

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

16. **If patient/encounter level validity testing was provided, was the method described and appropriate for assessing the accuracy of ALL critical data elements? NOTE:** Data element validation from the literature is acceptable.

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (patient/encounter level testing was not performed)

17. **Method of establishing validity at the accountable-entity level:**

    ☒ **Face validity**

    ☒ **Empirical validity testing at the accountable-entity level**

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

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

    ☒ **Yes**

    ☐ **No**

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

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

    - Empirical validity testing was conducted using Medicaid Analytic eXtract (MAX) data collected by the Centers for Medicare and Medicaid Services (CMS).
    - The MAX data was used to perform cross-state comparisons of TCD screenings among children with sickle cell anemia.
    - The MAX data was used as CMS develops and maintains these data for public use using administrative claims submitted by each state's Medicaid program.
    - Empirical testing compared the MAX data for the state of Michigan to the Michigan Medicaid data from 2007-2009.
    - Face validity was established by a panel of national experts and advocates for families of children with sickle cell disease (SCD) convened by the Quality Measurement, Evaluation, Testing, Review, and Implementation Consortium (Q-METRIC). The panel assessed whether the performance of the measure would result in improved quality of care for children with SCD, specifically in respect to TCD screening.
    - The panel weighed evidence to determine if performance of TCD would improve the quality of care provided to patients.

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

    - Empirical validity testing resulted in the number of TCD cases among children with sickle cell anemia ranging from 45 to 114 screenings in the claims acquired directly from the Medicaid data warehouse, versus a range of 26 to 93 screenings from MAX data for the same time period.
    - Face validity was rated as very high by the Q-METRIC expert panel. Receiving a rating of 8.5 out of 9, it was concluded that the performance of TCD, as outlined in the measure, would improve the quality of care provided to patients and distinguish good from poor quality.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    - The developer did not list any exclusions.

22. **Risk Adjustment**

22a. **Risk-adjustment method**

☒ None (only answer Question 20b and 20e) ☐ Statistical model ☐ Stratification

☐ Other method assessing risk factors (please specify)

22b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

22c. **Social risk adjustment:**

22c.1 Are social risk factors included in risk model? ☐ Yes ☐ No ☒ Not applicable

22c.2 Conceptual rationale for social risk factors included? ☐ Yes ☐ No

22c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☐ No

22d.**Risk adjustment summary:**

22d.1 All of the risk-adjustment variables present at the start of care? ☐ Yes ☐ No

22d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

22d.3 Is the risk adjustment approach appropriately developed and assessed? ☐ Yes ☐ No

22d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☐ Yes ☐ No

22d.5.Appropriate risk-adjustment strategy included in the measure? ☐ Yes ☐ No

22e**. Assess the risk-adjustment approach**

- No justification provided for not risk adjusting.

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

For cost/resource use measures, does this measure identify meaningful differences about cost and resource use between the measured entities?

- No concerns regarding identifying meaningful differences.

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

- No concerns as only one data source were used.

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

- Given the use of administrative data, non-responses were not tested. The developer did not anticipate the missing data to be different across groups.

26. **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 accountable-entity level testing has been conducted)

☒ **Moderate** (NOTE: Moderate is the highest eligible rating if accountable-entity 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 accountable-entity level and the patient/encounter level is required; if not conducted, should rate as INSUFFICIENT.)

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

    Threats to validity empirically assessed (Box 1) -> Empirical validity testing conducted using the measure as specified (Box 2) -> Empirical validity conducted at the accountable entity level (Box 5) -> Validity testing method described and appropriate (Box 6) -> Moderate certainty or confidence (Box 7b) -> Moderate rating

ADDITIONAL RECOMMENDATIONS

28. **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.**
    - Further discussion regarding not risk adjusting can be held by the Standing Committee.

Criteria 1: 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

---

1ma.01. Indicate whether there is new evidence about the measure since the most recent maintenance evaluation. If yes, please briefly summarize the new evidence, and ensure you have updated entries in the Evidence section as needed.

[Response Begins]

No

[Response Ends]

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

2021 Submission:

Updated evidence information here.

2018 Submission:

Evidence from the previous submission here.

1a. Evidence

1a.01. Provide a logic model.

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.

[Response Begins]

Transcranial Doppler (TCD) ultrasonography measures the blood flow velocity in cerebral arteries, specifically the distal internal carotid artery and the proximal middle cerebral artery. High blood velocities are indicative of an upcoming stroke and the need to begin stroke prevention efforts among children with sickle cell anemia. Stroke prevention efforts result in a substantial reduction in the incidence of stroke among children with sickle cell anemia.

[Response Ends]

1a.02. Select the type of source for 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.

[Response Begins]

Clinical Practice Guideline recommendation (with evidence review)

[Response Ends]

If the evidence is not based on a systematic review, skip to the end of the section and do not complete the repeatable question group below. If you wish to include more than one systematic review, add additional tables by clicking "Add" after the final question in the group.

Evidence - Systematic Reviews Table (Repeatable)

Group 1 - Evidence - Systematic Reviews Table

1a.03. Provide the title, author, date, citation (including page number) and URL for the systematic review.

[Response Begins]

National Heart, Lung, and Blood Institute. Evidence Based Management of Sickle Cell Disease. 2014; http://www.nhlbi.nih.gov/health-pro/guidelines/sickle-cell-disease-guidelines/sickle-cell-disease-report.pdf

[Response Ends]

1a.04. Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the systematic review.

[Response Begins]

Recommendations

1. In children with SCA, screen annually with TCD according to methods employed in the STOP studies, beginning at age 2 and continuing until at least age 16. **(Strong Recommendation, Moderate-Quality Evidence)**
2. In children with conditional (170-199 cm/sec) or elevated (>200 cm/sec) TCD results, refer to a specialist with expertise in chronic transfusion therapy aimed at preventing stroke. **(Strong Recommendation, High-Quality Evidence)**
3. In children with genotypes other than SCA (e.g., HbSβ+-thalassemia or HbSC), do not perform screening with TCD. **(Strong Recommendation, Low-Quality Evidence)**
4. In asymptomatic children with SCD, do not perform screening with MRI or CT. **(Moderate Recommendation, Low-Quality Evidence)**
5. In asymptomatic adults with SCD, do not perform screening with neuroimaging (TCD, MIR, or CT). **(Moderate Recommendation, Very Low-Quality Evidence)**

[Response Ends]

1a.05. Provide the grade assigned to the evidence associated with the recommendation, and include the definition of the grade.

[Response Begins]

Strong Recommendation, Moderate-Quality Evidence

Grade of Recommendation	Clarity of Risk/Benefit	Quality of Supporting Evidence	Implications
Strong recommendation Moderate-quality evidence	Benefits clearly outweigh harms and burdens, or vice versa	Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise evidence), or unusually strong evidence from unbiased observational studies	Recommendation can apply to most patients in most circumstances. Further research (if performed) is likely to have an impact on our confidence in the estimate of effect and may change the estimate.

[Response Ends]

1a.06. Provide all other grades and definitions from the evidence grading system.

[Response Begins]

Grade Recommendations – A Closer Look

Grade of Recommendation	Clarity of Risk/Benefit	Quality of Supporting Evidence	Implications
Strong recommendation High-quality evidence	Benefits clearly outweigh harms and burdens, or vice versa	Consistent evidence from well-performed RCTs or exceptionally strong evidence from unbiased observational studies	Recommendation can apply to most patients in most circumstances. Further research is very unlikely to change our confidence in the estimate of effect.
Strong recommendation Moderate-quality evidence	Benefits clearly outweigh harms and burdens, or vice versa	Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise evidence), or unusually strong evidence from unbiased observational studies	Recommendation can apply to most patients in most circumstances. Further research (if performed) is likely to have an impact on our confidence in the estimate of effect and may change the estimate.
Strong recommendation Low-quality evidence	Benefits clearly outweigh harms and burdens, or vice versa	Evidence for at least one critical outcome from observational studies, from RCTs with serious flaws, or indirect evidence	Recommendation may change when higher quality evidence becomes available. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and is like to change the estimate.

Grade of Recommendation	Clarity of Risk/Benefit	Quality of Supporting Evidence	Implications
Strong recommendation Very low-quality evidence (very rarely applicable)	Benefits clearly outweigh harms and burdens, or vice versa	Evidence for at least one of the critical outcomes from unsystematic clinical observations or very indirect evidence	Recommendation may change when higher quality evidence becomes available; any estimate of effect, for at least one critical outcome, is very uncertain.
Weak recommendation High-quality evidence	Benefits closely balanced with harms and burdens	Consistent evidence from well-performed RCTs or exceptionally strong evidence from unbiased observational studies	The best action may differ depending on circumstances or patient or societal values. Further research is very unlikely to change our confidence in the estimate of effect.

[Response Ends]

1a.07. Provide the grade assigned to the recommendation, with definition of the grade.

[Response Begins]

See above

[Response Ends]

1a.08. Provide all other grades and definitions from the recommendation grading system.

[Response Begins]

See above

[Response Ends]

1a.09. Detail the quantity (how many studies) and quality (the type of studies) of the evidence.

[Response Begins]

A total of 2 randomized control trials (RCTs) and 50 observational studies are included in the body of evidence (8 retrospective observational studies, 23 prospective observational studies, 18 cross-sectional studies, and 1 case series).

[Response Ends]

1a.10. Provide the estimates of benefit, and consistency across studies.

[Response Begins]

Receipt of TCD screening does not directly impact the risk of stroke among children with sickle cell anemia; however, the indication of high-risk of stroke obtained from TCD screening (blood flow velocity>200cm/sec) prompts the initiation of primary stroke prevention efforts in the form of blood transfusions. For brevity, we have included estimates of benefit

and consistency among studies within the body of evidence directly related to the process of TCD screening and the health-related outcome of primary stroke prevention among children with sickle cell anemia.

The majority of the studies used a standard definition of an abnormal TCD screening result (blood flow velocity>200cm/sec). A handful of studies used a looser definition, classifying velocities of over 170cm/sec as abnormal; however, these children would have been included in the definition of conditional TCD screening result in the other studies. Studies reported between 2% and 33% abnormal TCD screening results within their study populations; this large range may be attributable to differing study population inclusion criteria

All studies investigating the relationship between blood flow velocity as detected by TCD screening and stroke risk show that children with high blood flow velocities in the cerebral vessels are at a significantly increased risk of stroke. Adams (1992) reported in a prospective observational study that among 7 children who had a stroke within the study period (overall n=190), 6 children had an abnormal TCD screening result (Fisher's exact p-value<0.00001). Adams (2004) also reported that among 2,342 children with SCD who received a TCD screen, risk of stroke with abnormal TCD was much higher than with normal results (p-value<.001), conditional findings (p-value<.001), or inadequate TCD results (p-value=.002).

All studies that assessed stroke rates pre- and post-TCD screening recommendations found a significantly decreased rate of first stroke among children with sickle cell anemia post-TCD recommendations when compared with the pre-TCD recommendation time period. Armstrong-Wells (2008) reported a stroke rate of 0.44 per 100 pre-TCD recommendations and a stroke rate of 0.19 per 100 person-years post-TCD recommendations; Enningful-Eghan (2010) reported a stroke rate of 0.67 per 100 person-years pre-TCD recommendations and a post-TCD stroke rate of 0.06 per 100 person-years (p-value<0.0001). In addition, McCarville (2008) showed significantly decreasing stroke rates with increasing TCD use (p-value=0.045).

[Response Ends]

1a.11. Indicate what, if any, harms were identified in the study.

[Response Begins]

No harm is expected through the receipt of TCD screening; therefore, there is no negative affect of TCD screening on the net benefit.

[Response Ends]

1a.12. Identify any new studies conducted since the systematic review, and indicate whether the new studies change the conclusions from the systematic review.

[Response Begins]

N/A

[Response Ends]

1a.13. If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, describe the evidence on which you are basing the performance measure.

[Response Begins]

[Response Ends]

1a.14. Briefly synthesize the evidence that supports the measure.

[Response Begins]

[Response Ends]

1a.15. Detail the process used to identify the evidence.

[Response Begins]

[Response Ends]

1a.16. Provide the citation(s) for the evidence.

[Response Begins]

[Response Ends]

1b. Performance Gap

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

Children with sickle cell anemia (HbSS) have over three hundred times the stroke risk than children with normal hemoglobin (Verduzco and Nathan, 2009). Without intervention, approximately 11% of children with sickle cell anemia will have a stroke by age 20 (Verduzco and Nathan, 2009; Ohene-Frempong et al., 1998). Transcranial Doppler (TCD) ultrasonography measures the blood velocities within the cerebral vessels (Adams et al., 1997; Adams et al., 1992). Children over the age of 2 with a time-average mean maximum blood flow velocity of 200cm/sec or greater as measured by TCD ultrasonography have been shown to have 27 times the risk of stroke than children with velocities less than 200cm/sec. This corresponds to a 40% risk of stroke among those with high velocities within 3 years (Adams et al., 1997). Initiation of chronic blood transfusions reduces the risk of stroke by 92% among children at highest risk of stroke as identified through TCD screening (Adams et al., 1997; Adams et al., 1992). TCD screening is a reasonable method to assess stroke risk among children with sickle cell anemia, as it is safe, non-invasive and low cost (Markus, 2000). Although other predictors of stroke have been examined, such as hematocrit levels and white blood cell count, TCD velocities have been shown to be the only independent predictor of stroke (Adams et al., 1992). Given the importance of TCD screening to stroke prevention among children with sickle cell anemia, the National Heart, Lung, and Blood Institute (NHLBI) recommends each child with sickle cell anemia receive one TCD screen per year from ages 2 to 16 years (National Heart, Lung, and Blood Institute, 2014). Although the benefits of TCD screening among children with sickle cell anemia have been known since the late nineties, prior studies indicate that TCD screening rates are low. However, these reports are limited in their generalizability, as they are often focused on a single healthcare provider or registry. This measure establishes a claims-based method for identifying receipt of TCD screening among larger and broader populations of children with sickle cell anemia. The measure specifications are reflective of the guidelines from the NHLBI, and the performance scores calculated through this measure will identify areas in need of improvement in receipt of TCD screening among children with sickle cell anemia.

Citations:

Adams RJ, McKie VC, Carl EM, et al. Long-term stroke risk in children with sickle cell disease screened with transcranial Doppler. Ann Neurol. Nov 1997;42(5):699-704.

Adams R, McKie V, Nichols F, et al. The use of transcranial ultrasonography to predict stroke in sickle cell disease. N Engl J Med. Feb 27 1992;326(9):605-610.

Markus HS. Transcranial Doppler ultrasound. Br Med Bull. 2000;56(2):378-388.

National Heart Lung and Blood Institute. Evidence Based Management of Sickle Cell Disease. 2014; http://www.nhlbi.nih.gov/health-pro/guidelines/sickle-cell-disease-guidelines/sickle-cell-disease-report.pdf. Accessed 11/11, 2014.

Ohene-Frempong K, Weiner SJ, Sleeper LA, et al. Cerebrovascular accidents in sickle cell disease: rates and risk factors. Blood. Jan 1 1998;91(1):288-294.

Verduzco LA, Nathan DG. Sickle cell disease and stroke. Blood. Dec 10 2009;114(25):5117-5125.

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and 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 (4b) under Usability and Use.

[Response Begins]

Rates of transcranial Doppler screening among children with sickle cell anemia enrolled in Medicaid, by state, 2005-2010

Florida

(Year: Numerator/Denominator = Rate)

2005: 113/526 = 21.5%

2006: 121/489 = 24.7%

2007: 133/449 = 29.6%

2008: 171/502 = 34.1%

2009: 264/697 = 37.9%

2010: 339/734 = 46.2%

Illinois

(Year: Numerator/Denominator = Rate)

2005: 65/250 = 26.0%

2006: 85/276 = 30.8%

2007: 70/278 = 25.2%

2008: 78/291 = 26.8%

2009: 90/338 = 26.6%

2010: 86/302 = 28.5%

Louisiana

(Year: Numerator/Denominator = Rate)

2005: 150/364 = 41.2%

2006: 141/321 = 43.9%

2007: 164/322 = 50.9%

2008: 167/334 = 50.0%

2009: 164/356 = 46.1%

2010: 168/361 = 46.5%

Michigan

(Year: Numerator/Denominator = Rate)

2005: 27/240 = 11.3%

2006: 35/219 = 16.0%

2007: 26/243 = 10.7%

2008: 49/228 = 21.5%

2009: 93/259 = 35.9%

2010: 104/240 = 43.3%

South Carolina

(Year: Numerator/Denominator = Rate)

2005: 41/214 = 19.2%

2006: 37/189 = 19.6%

2007: 41/173 = 23.7%

2008: 48/124 = 38.7%

2009: 38/102 = 37.3%

2010: 68/134 = 50.7%

Texas

(Year: Numerator/Denominator = Rate)

2005: 18/258 = 7.0%

2006: 15/292 = 5.1%

2007: 56/343 = 16.3%

2008: 89/352 = 25.3%

2009: 140/372 = 37.6%

2010: 146/370 = 39.5%

Total

(Year: Numerator/Denominator = Rate)

2005: 414/1852 = 22.4%

2006: 434/1786 = 24.3%

2007: 301/1326 = 22.7%

2008: 313/1297 = 24.1%

2009: 357/1359 = 26.3%

2010: 431/1329 = 32.4%

Michigan Medicaid, 2011-2018:

Year	Numerator_case_count	Denominator_case_count	TCD_rate
2011	**109**	**237**	**45.99%**
2012	**96**	**231**	**41.56%**

Year	Numerator_case_count	Denominator_case_count	TCD_rate
2013	108	229	47.16%
2014	110	238	46.22%
2015	102	257	39.69%
2016	128	345	37.10%
2017	107	301	35.55%
2018	134	312	42.95%
Overall	894	2150	41.58%

New York State Medicaid, 2011-2018

Year	TCD num	TCD denom	TCD rate
2011	323	818	0.39
2012	362	899	0.4
2013	413	932	0.44
2014	372	893	0.42
2015	416	1021	0.41
2016	432	1106	0.39
2017	461	1095	0.42
2018	497	1097	0.45
Overall	3276	7861	0.42

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, 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. Include citations.

[Response Begins]

N/A

[Response Ends]

1b.04. 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.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. 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 (4b) under Usability and Use.

[Response Begins]

There are no gender disparities in TCD screening among children with sickle cell anemia (chi-square=1.2, p-value=0.28). The data used for performance scores is state Medicaid programs; therefore, there are no disparities identified by insurance or socioeconomic status. Younger children (ages 2-6) were more likely to receive TCD screening than older children (chi-square=99.01, p-value<0.0001). For those 2 to 6 years old, 36% received a TCD screen; for those ages 7 to 11 years, 31% received a TCD screen; and for those ages 12-15 years, 25% were screened.

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, 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 above.

[Response Begins]

N/A

[Response Ends]

Criteria 2: 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.

---

spma.01. Indicate whether there are changes to the specifications since the last updates/submission. If yes, update the specifications in the Measure Specifications section of the Measure Submission Form, and explain your reasoning for the changes below.

[Response Begins]

No

[Response Ends]

spma.02. Briefly describe any important changes to the measure specifications since the last measure update and provide a rationale.

For annual updates, please explain how the change in specifications affects the measure results. If a material change in specification is identified, data from re-testing of the measure with the new specifications is required for early maintenance review.

For example, specifications may have been updated based on suggestions from a previous NQF CDP review.

[Response Begins]

N/A

[Response Ends]

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]

Transcranial Doppler Ultrasonography Screening Among Children with Sickle Cell Anemia

[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]

The percentage of children ages 2 through 15 years old with sickle cell anemia (Hemoglobin SS) who received at least one transcranial Doppler (TCD) screening within a year.

[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]

Neurology: Stroke/Transient Ischemic Attack (TIA)

[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]

Disparities Sensitive

Screening

[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

Children (Age < 18)

[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Health Plan

[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.

[Response Begins]

Other

Outpatient Services

[Response Ends]

sp.09. 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. If no URL is available, indicate "none available".

[Response Begins]

https://chear.org/sites/default/files/TranscranialDopplerScreeningMeasureSpecification.pdf

[Response Ends]

sp.11. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff . Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

[Response Begins]

Available in attached Excel or csv file

[Response Ends]

Attachment: 2797_SCA_TCD Measure Appendix Tables 20180501.xlsx

sp.12. State the numerator.

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.

[Response Begins]

The numerator is the number of children ages 2 through 15 years old with sickle cell anemia who received at least one TCD screening within the measurement year.

[Response Ends]

sp.13. Provide details needed to calculate the numerator.

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 sp.11.

[Response Begins]

Cases from target population with target process (Receipt of TCD screening): Receipt of TCD screening is identified as the presence of at least one CPT code for any of five acceptable ultrasonography tests within the measurement year among children in the target population. Acceptable CPT codes are: 93886 (complete study), 93888 (limited study), 93890 (vasoreactivity study), 93892 (emboli detection without intravenous microbubble injection), and 93893 (emboli detection with intravenous microbubble injection).

[Response Ends]

sp.14. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

The denominator is the number of children ages 2 through 15 years with sickle cell anemia within the measurement year.

[Response Ends]

sp.15. Provide details needed to calculate the denominator.

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 sp.11.

[Response Begins]

Children with sickle cell anemia are identified through the presence of at least three separate healthcare encounters related to sickle cell anemia (defined as hemoglobin [Hb]SS) within the measurement year. Sickle cell anemia-related healthcare encounters are identified through ICD codes. The ICD-9-CM codes to identify HbSS-related healthcare encounters are as follows: 282.61 (Hb-SS disease w/o crisis) and 282.62 (Hb-SS disease with crisis). The ICD-10-CM codes for HbSS-related healthcare encounters are as follows: D57.1 (Hb-SS disease without crisis), D57.00 (Hb-SS disease with crisis, unspecified); D57.01 (Hb-SS disease with acute chest syndrome); and D57.02 (Hb-SS disease with splenic sequestration). Children ages 2 through 15 years are included within the target population (i.e., must not have a 2nd or 16th birthday within the measurement year).

It is important to note that accurate calculation of this measure requires that the target population be selected from among children who have all of their health services for the measurement year included in the administrative claims data set. For children who have dual enrollment in other health plans, their claims may not be complete since some of their health services may have been paid for by another health plan. Inclusion of children with other health insurance would potentially cause this measure to be understated. As a consequence, this measure requires that children must not only be continuously enrolled within the health plan from which claims are available, the enrollment files must also be assessed to determine whether other forms of health insurance existed during the measurement year. Children with evidence of other insurance during the measurement year (i.e., coordination of benefits) are excluded from the target population.

[Response Ends]

sp.16. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

There are no denominator exclusions.

[Response Ends]

sp.17. Provide details needed to calculate the denominator exclusions.

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 sp.11.

[Response Begins]

N/A

[Response Ends]

sp.18. Provide all information required to stratify the measure results, if necessary.

Include 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 in the Data Dictionary field.

[Response Begins]

N/A

[Response Ends]

sp.19. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

[Response Begins]

No risk adjustment or risk stratification

[Response Ends]

sp.20. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

sp.21. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score

[Response Begins]

Better quality = Higher score

[Response Ends]

sp.22. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]

1. Identify the denominator: Determine the eligible population using administrative claims. The eligible population is all individuals who satisfy all specified criteria, including age, continuous enrollment, and diagnosis requirements within the measurement year.

2. Identify the numerator: Identify numerator events using administrative claims for all individuals in the eligible population (denominator) within the measurement year.

3. Calculate the rate (numerator / denominator).

[Response Ends]

sp.25. If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

[Response Begins]

This measure does not involve sampling; all sickle cell anemia cases meeting the inclusion criteria are included.

sp.28. Select only the data sources for which the measure is specified.

[Response Begins]

Claims

[Response Ends]

sp.29. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]

N/A

[Response Ends]

sp.30. Provide the data collection instrument.

[Response Begins]

No data collection instrument provided

[Response Ends]

2ma.01. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Reliability - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

No

[Response Ends]

2ma.02. Indicate whether additional empirical reliability testing at the accountable entity level has been conducted. If yes, please provide results in the following section, Scientific Acceptability: Validity - Testing. Include information on all testing conducted (prior testing as well as any new testing).

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous Submission:

Testing from the previous submission here.

[Response Begins]

Yes

[Response Ends]

2ma.03. For outcome, patient-reported outcome, resource use, cost, and some process measures, risk adjustment/stratification may be conducted. Did you perform a risk adjustment or stratification analysis?

[Response Begins]

No

[Response Ends]

2ma.04. For maintenance measures in which risk adjustment/stratification has been performed, indicate whether additional risk adjustment testing has been conducted since the most recent maintenance evaluation. This may include updates to the risk adjustment analysis with additional clinical, demographic, and social risk factors.

Please update the Scientific Acceptability: Validity - Other Threats to Validity section.

Note: This section must be updated even if social risk factors are not included in the risk adjustment strategy.

[Response Begins]

No additional risk adjustment analysis included

[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.

- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the

Submitting Standards webpage

- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the

2021 Measure Evaluation Criteria and Guidance .

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

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

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient

preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

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

OR

• rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

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

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions

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

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

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

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

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

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one

percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Scientific Acceptability sections. For example:

2021 Submission:

Updated testing information here.

2018 Submission:

Testing from the previous submission here.

2a. Reliability

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]

Claims

[Response Ends]

2a.02. 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).

[Response Begins]

2021 Submission:

- Michigan Medicaid administrative claims data provided by the Michigan Department of Health and Human Services (MDHHS)
- New York State Medicaid administrative data provided by the New York State Department of Health
- Medicaid Analytic eXtract (MAX) administrative claims data for six state Medicaid programs provided by the Centers for Medicare & Medicaid Services (CMS)
- Michigan Newborn Screening (NBS) Results
- New York State Newborn Screening (NBS) Results
- Medical record data from Children's Hospital of Michigan (CHM), Detroit, Michigan; Hurley Medical Center (HMC), Flint, Michigan; and University of Michigan Health System (UMHS), Ann Arbor, Michigan

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]

2021 Submission:

Michigan Medicaid data 2007-2011, 2016; New York State Medicaid data: 2016; MAX data: 2005-2012; Michigan Newborn Screening: 1987-2014; New York State Newborn screening: 2006-2013; Medical record data: 2012

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Health Plan

[Response Ends]

2a.05. List the measured entities 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.

[Response Begins]

2021 Submission:

- **Michigan Medicaid:** all Medicaid claims for children ages 3 months through 5 years continuously enrolled for at least one year in Michigan Medicaid in 2007-2010; all Medicaid claims for children ages 1 through 17 continuously enrolled for at least one year in 2010, 2011, or **2016.**
- **New York State Medicaid: all Medicaid claims for children ages 1 through 17 continuously enrolled in New York Medicaid in 2016.**
- **MAX data:** all Medicaid claims reported to CMS for children ages 3 months through 5 years continuously enrolled for at least one year from 2005**-2012** within one of the six following state Medicaid programs with moderate to high prevalence of sickle cell anemia: Florida, Illinois, Louisiana, Michigan, South Carolina, and Texas.
- **Michigan NBS data** consisted of all births within the state of Michigan from 1987-**2014.**
- ***New York State NBS data* consisted of all sickle cell births within the state of New York from 2006-2013.**
- **Medical record data** were obtained from three hospitals in 2012: CHM, HMC, and UMHS. These three large medical centers are located in urban areas in Michigan, reflective of the residence of the vast majority (~70%) of children with sickle cell anemia living in Michigan:
    - CHM is a tertiary medical center located in Detroit, Michigan.
    - HMC is a tertiary medical center located in Flint, Michigan.
    - UMHS is an academic medical center located in Ann Arbor, Michigan.

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

2021 Submission:

- **Michigan Medicaid** data for ICD-9-CM denominator validation (2010 and 2011) included 865 children in 2010 and 863 children in 2011. These children were included if continuously enrolled in either 2010 or 2011, with at least one sickle cell disease-related administrative claim, and a newborn screening result available from 1987-2010.
- ***Michigan Medicaid*** **data for ICD-10-CM denominator validation (2016 only) included 1,597 children with at least one D57x diagnosis code in 2016.**
- ***Michigan Medicaid*** data for performance score validation included 114 children in 2007, 118 children in 2008, 149 children in 2009, and 141 children in 2010. These children were included if continuously enrolled in Michigan Medicaid for at least one year from 2007-2010, and met the ICD-9-CM case definition for sickle cell anemia.
- ***New York State Medicaid*** **data included 2,454 children with at least one D57x diagnosis code in 2016.**
- ***MAX data*** (2005-**2012)** included 2,821 children enrolled for a total of 5,014 person-years. The breakdown of person-years by state is as follows: Florida: 1,619; Louisiana: 855; Texas: 897; Illinois: 622; Michigan: 580; South Carolina: 441. Children were included if they met the denominator criteria for sickle cell anemia.
- ***Michigan NBS data f*****or ICD-10-CM validation included a total of 1,457 children that had both a D57x diagnosis code in 2016 and a Newborn Screening Result available from 1997-2014.**
- ***New York State NBS data*** **included 2,454 children with a D57x diagnosis code in 2016 and born between 2006-2013.**
- ***Medical record data*** included 34 children with sickle cell anemia ages 3 months to 5 years who were enrolled in Michigan Medicaid and received care at one of three sickle cell centers in Michigan (CHM, HMC, UMHS) during 2012.

Across all datasets, the majority of children were Black, with a similar proportion of males and females.

[Response Ends]

[Response Begins]

2021 Submission:

- Reliability testing data: MAX
- Validity testing data: Michigan Medicaid, MAX, Michigan Newborn Screening, **New York State Medicaid, New York State Newborn Screening,** and medical records
- Identification of meaningful differences in performance: MAX data

[Response Ends]

2a.08. List 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.

[Response Begins]

The data do not include data on patient-level social risk factors; however, all children included in the data were enrolled in Medicaid, which is a proxy for low socioeconomic status.

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.07 check patient or encounter-level data; in 2a.08 enter "see validity testing section of data elements"; and enter "N/A" for 2a.09 and 2a.10.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.

[Response Begins]

Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)

[Response Ends]

2a.10. For each level of reliability testing 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.

[Response Begins]

Initial Submission:

The reliability of MAX data to evaluate TCD screening is of high importance since this is the only national source of state Medicaid data available upon which state-to-state comparisons may be conducted. The reliability of this measure was calculated using a signal-to-noise analysis. The signal-to-noise analysis was focused on assessing the reliability to confidently distinguish the performance of one state's Medicaid program from that of another state. For this approach, reliability was estimated with a beta-binomial model (RAND Corporation, TR-653-NCQA, 2009).

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]

Initial Submission:

State-specific reliability results showed that the reliability based on signal-to-noise analysis ranged from 0.96 to 0.99, with a median of 0.98.

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]

Initial Submission:

State-specific reliability is very good; observed reliability was consistently greater than 0.95. In general, reliability scores can range from 0.0 (all variation is attributable to measurement error) to 1.0 (all variation is caused by real differences). While there is not a clear cut-off for minimum reliability level, values above 0.7 are considered sufficient to distinguish differences between some states and the mean; reliability values above 0.9 are considered sufficient to see differences between states (RAND Corporation, TR-653-NCQA, 2009). The median reliability observed across states was 0.98 (range: 0.96-0.99), which is consistent with a high degree of reliability.

[Response Ends]

2b. Validity

2b.01. Select the level of validity testing that was conducted.

[Response Begins]

Patient or Encounter-Level (data element validity must address ALL critical data elements)

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)

[Response Ends]

2b.02. 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.

[Response Begins]

Critical Data Elements

2021 Submission

Denominator: The denominator (children with sickle cell anemia) has been validated using both ICD-9-CM **and ICD-10-CM diagnosis codes.** The process for each separate validation (ICD-9-CM and ICD-10-CM) is detailed below:

ICD-9-CM denominator validation: The accuracy of the case definition (at least three claims for sickle cell anemia [Hb SS] within the measurement year) to identify children with sickle cell anemia was assessed through comparison with the gold

standard of newborn screening results for the state of Michigan for children enrolled in Michigan Medicaid in 2010 and 2011 with at least one SCD-related healthcare claim within their enrollment year(s). The area under the receiver operating characteristic (ROC) curve, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for the case definition. As a comparison, these values were also calculated for those with a minimum of at least one or two Hb SS claims within each year.

ICD-10-CM denominator validation: **We developed, tested, and validated the performance of ICD-10-CM case definitions for SCA through the following 4-step process: (a) developing candidate SCA case definitions; (b) identifying a test population for the definitions; (c) testing the accuracy of candidate case definitions; and (d) testing case definitions within an independent population. Briefly, using specific SCA-related (D5700, D5701, D5702) and non-specific (D571) diagnosis codes, 23 SCA case definitions were applied to Michigan Medicaid claims (2016) to identify children with SCA. Measures of performance (sensitivity, specificity, area under ROC curve) were calculated using newborn screening results as the gold standard. A parallel analysis was conducted using New York State Medicaid claims and newborn screening data.**

Additional details regarding the methodology of validating the ICD-10-CM definition can be found at: Reeves SL, Madden B, Wu M, et al. Performance of ICD-10-CM diagnosis codes for identifying children with Sickle Cell Anemia. Health Serv Res. 2020;55(2):310-317. doi:10.1111/1475-6773.13257

Initial Submission

Numerator: The accuracy of administrative claims in identifying receipt of TCD screening was assessed through comparison to the gold standard of medical charts. An audit was conducted by trained medical record abstractors to compare administrative claims data with corresponding medical records data. Medical records were abstracted for all children meeting the TCD screening measure specification criteria; agreement between the medical records and the administrative claims was assessed using kappa. Furthermore, the sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) of administrative claims for receipt of TCD screening were calculated; the medical charts were the gold standard for comparison. In addition, the reliability of the data element abstracted from the medical chart was assessed by identifying a subset of the charts to be re-abstracted by another trained medical record abstractor; the results of the two abstractors were compared using percent agreement and kappa.

Empirical Validity Testing of Performance Measure

Although a state would typically have direct access to its own Medicaid data, it is unlikely that a state would have similar access to other states' data for comparison. However, CMS develops and maintains standardized Medicaid Analytic eXtract (MAX) data for public use using administrative claims submitted by each state Medicaid program. The MAX data are the only national, person-level administrative claims dataset available for the Medicaid program. As a consequence, MAX data, rather than data acquired directly from individual Medicaid programs, are likely to be used to perform cross-state comparisons of TCD screening among children with sickle cell anemia. Since states submit their Medicaid data to CMS for conversion into the MAX datasets, a state's own Medicaid data can be considered the authoritative source for administrative claims.

Our empirical validity testing of this performance measure compared the MAX data for the state of Michigan (obtained from CMS) to the gold standard of Michigan Medicaid data (obtained directly from Michigan's claims data warehouse) for the same time period (2007-2009). Note that the testing time period was constrained to align with the most recent MAX data available from CMS at the time of this analysis. Rates of TCD screening using each source of data were calculated and compared using z-tests for two proportions; for these tests, the null hypothesis was that the rate in each year would be the same in both Michigan Medicaid data and MAX data. Additionally, the correlation coefficient and squared correlation coefficient were calculated to identify the extent of the linear relationship between the two data sources.

Face Validity of Performance Measure Score

The face validity of this measure was established by a panel of national experts and advocates for families of children with SCD convened by the Quality Measurement, Evaluation, Testing, Review, and Implementation Consortium (Q-METRIC). The Q-METRIC expert panel included nationally recognized experts in SCD, representing hematology, pediatrics, and SCD family advocacy. In addition, measure validity was considered by experts in state Medicaid program operations, health plan quality measurement, health informatics, and health care quality measurement. In total, the Q-METRIC SCD

panel included 14 experts, providing a comprehensive perspective on SCD management and the measurement of quality metrics for states and health plans. The expert panel assessed whether the performance of the measure would result in improved quality of care for children with sickle cell disease. Specifically, in respect to TCD screening, the panel weighed evidence to determine if the performance of TCD as outlined in the measure would improve the quality of care provided to patients. The voting process to prioritize the measure was based on the ability of the measure to distinguish good from poor quality.

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]

2021 Submission:

Denominator: The denominator has been validated using both ICD-9-CM **and ICD-10-CM diagnosis codes.** The results for each separate validation (ICD-9-CM and ICD-10-CM) are detailed below:

ICD-9-CM denominator validation: In 2010, a case definition of three Hb SS (282.61, 282.62) claims within the year was 91.4% sensitive and 80% specific in identifying children with sickle cell anemia (Hb SS) (PPV: 80.4%; NPV: 91.3%). These results were replicated with nearly identical precision among the study population in 2011. In comparison, using a case definition of at least one Hb SS claim or at least two Hb SS claims to identify the study population resulted in substantially less specificity.

ICD-10-CM denominator validation: **Across the 23 case definitions, measures of performance varied, with sensitivities from 0.02-0.97 and specificities from 0.88-1.0. The case definition of ≥1 outpatient visit with a SCA-related (D5700, D5701, D5702) or D571 code had the highest area under the ROC curve, with a sensitivity of 95% and specificity of 92%. The same definition also had the highest performance in New York Medicaid (n=2,454), with a sensitivity of 94% and specificity of 86%.**

Initial Submission:

Numerator: For this comparison, 91 children with sickle cell anemia who were enrolled within Michigan Medicaid were successfully matched with their Michigan Medicaid administrative claims data. Among these children, TCD screening was identified in both the administrative claims data and the medical record review for 47 (51.6%) cases. Similarly, 41 (45.1%) cases were classified as not having a TCD in both data sources, yielding an overall agreement of 96.7% (kappa = 0.93, 95% confidence interval (CI): 0.86, 1). Using administrative claims to identify receipt of TCD screening resulted in a sensitivity of 94% (95% CI: 83%-99%), a specificity of 100% (95% CI: 91%-100%), a NPV of 93% (95% CI: 81%-99%), and a PPV of 93% (95% CI: 92%-100%) compared with the gold standard of medical records. Ten charts were also chosen for exploration of inter-rater reliability; the two trained abstractors had 100% agreement with each other for abstracting receipt of TCD screening from the medical records, resulting in a kappa of 1.00.

Empirical Validity Testing of Performance Measure

We compared rates of TCD screening from the gold standard of Michigan Medicaid data as compared to MAX data. The number of TCD cases among children with sickle cell anemia ranged from 45 to 114 screenings in the claims acquired directly from the Medicaid data warehouse, versus a range of 26 to 93 screenings from MAX data for the same time period.

Face Validity of Performance Measure Score

The Q-METRIC expert panel concluded that this measure has a very high degree of face validity through a detailed review of concepts and metrics considered to be essential to effective SCD management and treatment. Concepts and draft measures were rated by this group for their relative importance. This measure was among the most highly rated, receiving an average score of 8.5 (with 9 as the highest possible score). In addition, the expert panel concluded that the

performance of TCD as outlined in this measure would improve the quality of care provided to patients, and the measure would be able to distinguish good from poor quality.

[Response Ends]

2b.04. Provide 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?)

[Response Begins]

2021 Submission

Denominator: Using both ICD-9-CM **and ICD-10-CM diagnosis codes**, children with sickle cell anemia can be identified with a high level of accuracy in administrative data.

Initial Submission

Numerator: A kappa of greater than .81 is considered almost perfect agreement (Landis and Koch, 1997). In addition, the sensitivity, specificity, NPV and PPV are high. Given this evidence, we believe the validity of administrative claims in assessing receipt of TCD screening is very high.

Empirical Validity Testing of Performance Measure

Our results suggest that, compared with the gold standard of Michigan Medicaid data, MAX data has a very high degree of validity. When TCD screening was assessed for the same state (Michigan) from these two data sources for the same time period (2007-2009), no differences in rates were observed (all p-values >0.20). Additionally, the high values of the correlation coefficient and the squared correlation coefficient indicate a high level of reliability. Correlation coefficients of greater than 0.70 indicate a strong positive linear relationship; therefore, our results suggest that compared with Michigan Medicaid data, MAX data is highly valid. The squared correlation coefficient value of 0.96 indicates that nearly 96% of the variability in the MAX data from CMS for the state of Michigan can be explained by variation in the data received directly from the Michigan Medicaid program. This finding indicates that the strength of the relationship between the two data sources is extremely strong.

Face Validity of Performance Measure Score

Given the high rating of the Q-METRIC expert panel, we feel this measure has a very high degree of face validity.

[Response Ends]

2b.05. 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 in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

Initial Submission

Using the MAX data, the proportion of children receiving annual TCD screening was calculated for each year in the study period (2005 - 2010). We examined differences in performance across the 6 years included within this dataset. Logistic regression was used to estimate the associations between each year and receipt of TCD screening, with 2005 used as the reference category. Generalized estimating equation (GEE) models with robust standard errors were used to account for the correlation among children. Odds ratios with 95% confidence intervals were used to assess the final associations. The presence of trends in TCD screening rates were also assessed over time using linear regression. For all models, regression diagnostics were performed to assess normality of error variances.

[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

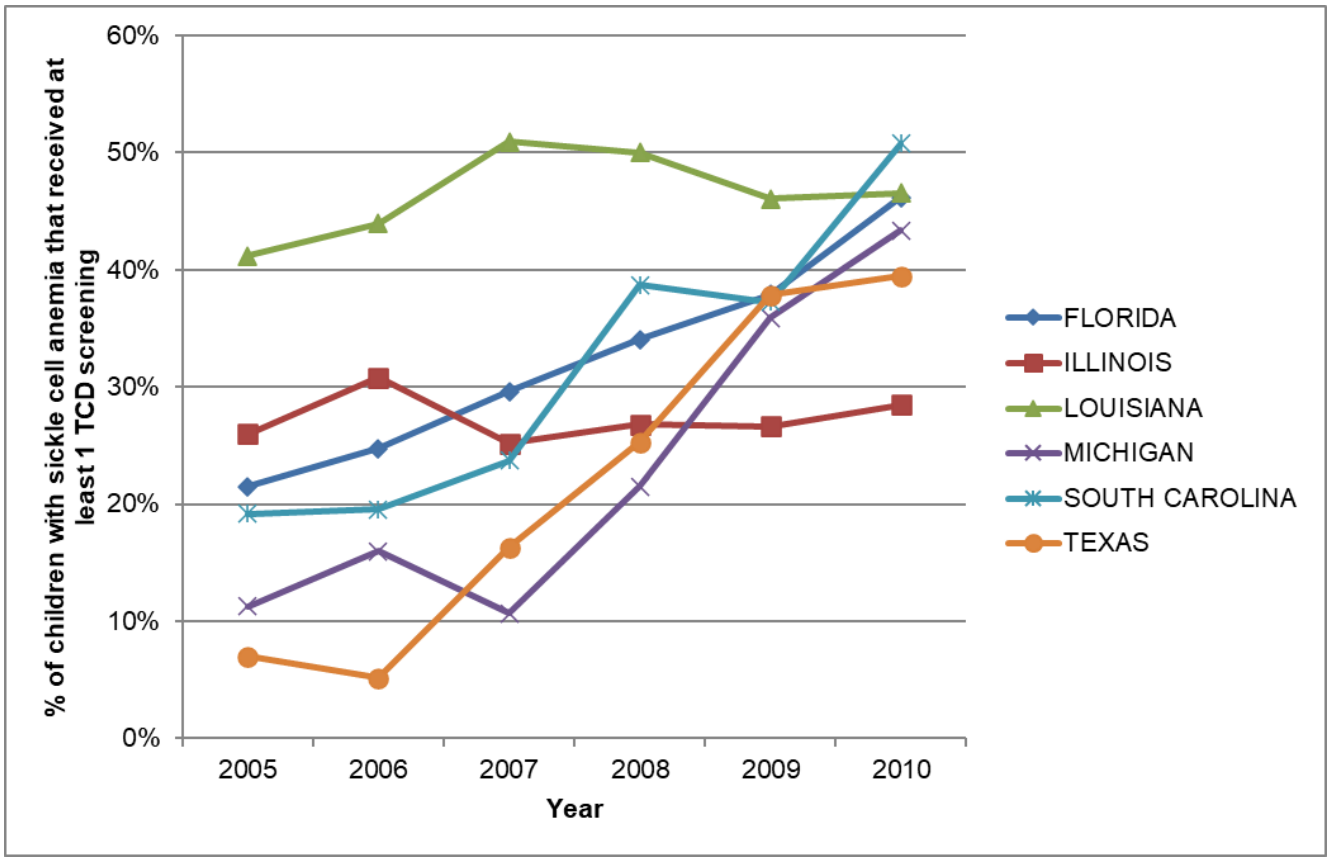
Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.

[Response Begins]

Initial Submission

The proportion of children receiving TCD screening ranged from 7% to 51% (Figure 1).

Figure 1. Trends for transcranial Doppler screening within the measurement year for children with sickle cell anemia, tested in six state Medicaid programs using MAX data, 2005-2010



Compared with 2005, children had higher odds of receiving TCD screening; these odds were statistically significant starting in 2007 (Table 3). Results from the linear regression model indicated that these rates did increase over time (p=0.0001).

Table 3. Odds of receipt of TCD screening among children with sickle cell anemia enrolled in 6 state Medicaid programs by year using MAX data, 2005-2010

Year	Odds Ratio	95% Confidence Interval	p-value
2005	Reference	Reference	N/A
2006	1.09	0.96, 1.25	0.17

Year	Odds Ratio	95% Confidence Interval	p-value
2007	1.26	1.10, 1.44	0.0008
2008	1.60	1.40, 1.83	<0.0001
2009	1.94	1.69, 2.22	<0.0001
2010	2.36	2.10, 2,70	<0.0001

[Response Ends]

2b.07. Provide 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.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]

Initial Submission

This measure was successfully able to distinguish differences in performance across years; the measure was also able to detect changes over time. As children in all years after 2005 had increased odds of receipt of TCD screening compared with children in 2005, these results demonstrate that the likelihood of receiving a TCD screening did increase significantly over time.

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]

Given that this measure uses administrative data, we did not test the extent of non-response. Although this measure is subject to all the usual limitations of administrative data, we do not anticipate that data would be missing differentially across any groups (e.g., missing data would not be systematically missing).

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]

As this measure uses administrative claims, we are unable to assess the frequency of missing data.

[Response Ends]

2b.10. Provide 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.

In other words, 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 was conducted, justify the selected approach for missing data.

[Response Begins]

As any data would not be missing differentially, we do not anticipate this would have a significant impact on our results.

[Response Ends]

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 eCQMs). 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.

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]

No, there is only one set of specifications for this measure

[Response Ends]

2b.12. 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. Indicate what statistical analysis was used.

[Response Begins]

[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]

[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]

[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]

N/A or no exclusions

[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

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?

[Response Begins]

N/A

[Response Ends]

2b.17. Provide 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.

[Response Begins]

N/A

[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, 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.

[Response Begins]

N/A

[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]

No risk adjustment or stratification

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]

[Response Ends]

2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]

N/A

[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]

[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and

within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]

[Response Ends]

2b.26. 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 control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]

[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]

N/A

[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]

[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]

[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]

[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]

[Response Ends]

Criteria 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.

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]

Coded by someone other than person obtaining original information (e.g., DRG, ICD-10 codes on claims)

[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

ALL data elements are in defined fields in electronic claims

[Response Ends]

3.03. 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 data elements not from electronic sources.

[Response Begins]

N/A

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

We have not developed an eCQM.

[Response Ends]

3.06. 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.

[Response Begins]

Availability of administrative claims data may vary by source (e.g., state Medicaid, national Medicaid). Although the measures are straightforward to implement in any administrative claims database, a knowledgeable programmer is required.

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail 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),

Attach the fee schedule here, if applicable.

[Response Begins]

N/A

[Response Ends]

Criteria 4: Use and Usability

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.

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

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 demonstrating performance improvement.

4a.01.

Check all current uses. For each current use checked, please provide:

Name of program and sponsor

URL

Purpose

Geographic area and number and percentage of accountable entities and patients included

Level of measurement and setting

[Response Begins]

Payment Program

Michigan Medicaid health plans are incentivized through Michigan Medicaid to improve rates of transcranial doppler screening among children with sickle cell anemia in southeast Michigan.

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Public reporting

Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

Quality Improvement (internal to the specific organization)

[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]

N/A

[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A 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.

[Response Begins]

N/A

[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail 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.

[Response Begins]

We have consulted with the Michigan Department of Health and Human Services as this measure is integrated into their Medicaid reporting systems.

We have provided assistance with interpretation of the measure specifications to other users interested in the measure, including University of Florida Health, Cincinnati Children's Hospital Medical Center, Centers for Medicare and Medicaid Services, health systems (Hurley Children's Hospital, University of Michigan), University of California San Francisco, New York Medicaid, Medical University of South Carolina, and Albert Einstein College of Medicine.

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]

Measure results are provided quarterly to Michigan Medicaid health plans.

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]

Feedback has indicated that users find the measure useful and valuable.

[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]

Feedback has indicated that users find the measure useful and valuable to understand the impact of quality improvement initiatives.

[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]

Feedback has indicated that users find the measure useful and valuable.

[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]

No measure modification has resulted from user feedback.

[Response Ends]

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, 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, provide an explanation. 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.

[Response Begins]

Michigan Medicaid health plans have recently begun quality improvement efforts related to this measure; therefore, improvements in measurement are not evident yet.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]

No unintended negative consequences to individuals or populations were identified during testing.

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]

N/A; although we anticipate the proportion of children with a hematologist visit will also increase as this measure improves.

[Response Ends]

Criteria 5: Related and 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.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)

[Response Begins]

3166: Antibiotic Prophylaxis Among Children with Sickle Cell Anemia

3595: Hydroxyurea Use Among Children with Sickle Cell Anemia

[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

N/A

[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

Yes

[Response Ends]

5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

[Response Begins]

N/A

[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]

There are no competing measures.

[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.: No appendix

Contact Information

Measure Steward (Intellectual Property Owner): Q-METRIC – The University of Michigan

Measure Steward Point of Contact: Reeves, Sarah, sleasure@umich.edu

Measure Developer if different from Measure Steward: Q-METRIC – The University of Michigan

Measure Developer Point(s) of Contact: Reeves, Sarah, sleasure@umich.edu

Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]

No appendix

[Response Ends]

2. List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

[Response Begins]

The face validity of this measure was established by a national panel of experts and advocates for families of children with sickle cell disease (SCD) convened by the Quality Measurement, Evaluation, Testing, Review, and Implementation Consortium (Q-METRIC) at the University of Michigan. The Q-METRIC Representative Panel included nationally recognized experts in SCD, representing hematology, pediatrics, and SCD family advocacy. The Q-METRIC Feasibility Panel included experts in state Medicaid program operations, health plan quality measurement, health informatics, and health care quality measurement. In total, the Q-METRIC SCD panels included 14 experts, providing a comprehensive perspective on SCD management and the measurement of quality metrics for states and health plans.

The Q-METRIC expert panels concluded that this measure has a very high degree of face validity through a detailed review of concepts and metrics considered to be essential to effective SCD management and treatment. Concepts and draft measures were rated by this group for their relative importance. This measure was among the most highly rated,

receiving an average score of 8.5 (with 9 as the highest possible score).

Sickle Cell Disease Representative Panel:
Samir Ballas, MD, Professor, Division of Hematology, Thomas Jefferson University, Philadelphia, PA
Mary E. Brown, President and Chief Executive Officer, Sickle Cell Disease Association, Los Angeles, CA
George Buchanan, MD, Pediatric Hematologist, University of Texas Southwest Medical Center at Dallas, TX
Peter Lane, MD, Pediatric Hematologist-Oncologist, Children's Healthcare of Atlanta Pediatric Hospital, Atlanta, GA
Suzette Oyeku, MD, Assistant Professor of Pediatrics, Albert Einstein College, Bronx, NY
Lynnie Reid, Parent Representative, Boston, MA
Elliott Vichinsky, MD, Pediatric Hematology-Oncology, Children's Hospital and Research Center, Oakland, CA
Winfred Wang, MD, Hematologist, St. Jude Children's Hospital, Memphis, TN

Sickle Cell Disease Feasibility Panel:
Cathy Call, BSN, MSC, Senior Policy Analyst and Director for Health Quality Research, Altarum Institute, Alexandria, VA
J. Mitchell Harris, PhD, Director Research and Statistics, Children's Hospital Association, (formerly NACHRI), Alexandria, VA
Kevin Johnson, MD, MS, Professor and Vice Chair of Biomedical Informatics, Vanderbilt University, Nashville, TN
Don Lighter, MD, MBA, FAAP, FACHE, Director, The Institute for Health Quality Research and Education, Knoxville, TN
Sue Moran, BSN, MPH, Director of the Bureau of Medicaid Program Operations and Quality Assurance, Michigan Department of Community Health, Lansing, MI
Joseph Singer, MD, Vice President Clinical Affairs, HealthCore, Inc., Wilmington, DE
C. Jason Wang, MD, PhD, Associate Professor of Pediatrics, Stanford School of Medicine, Stanford, CA

Q-METRIC Investigators:
Kevin J. Dombkowski, DrPH, MS, Research Associate Professor of Pediatrics, School of Medicine, University of Michigan, Ann Arbor, MI
Gary L. Freed, MD, MPH, Professor of Pediatrics, School of Medicine and Professor of Health Management and Policy, School of Public Health, University of Michigan, Ann Arbor, MI (principal investigator)

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]

2016

[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]

May 2019

[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]

Quarterly

[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]

April 2022

[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".

[Response Begins]

N/A

[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate "N/A".

[Response Begins]

This work was previously funded by the Agency for Healthcare Research and Quality (AHRQ) and the Centers for Medicare & Medicaid Services (CMS) under the CHIPRA Pediatric Quality Measures Program Centers of Excellence grant number U18 HS020516. AHRQ, in accordance to CHIPRA 42 U.S.C. Section 1139A(b), and consistent with AHRQ´s mandate to disseminate research results, 42 U.S.C. Section 299c-3, has a worldwide irrevocable license to use and permit others to use products and materials from the grant for government purposes, which may include making the materials available for verification or replication by other researchers and making them available to the health care community and the public, if such distribution would significantly increase access to a product and thereby produce substantial or valuable public health benefits.

Gary L. Freed, MD, MPH

Percy and Mary Murphy Professor of Pediatrics, School of Medicine

Professor of Health Management and Policy, School of Public Health

Principal Investigator, Q-METRIC

Child Health and Evaluation Research (CHEAR) Unit

Division of General Pediatrics

University of Michigan Hospital and Health Systems

Ann Arbor, MI 48109-5456

[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate “N/A”.

[Response Begins]

N

[Response Ends]