MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 2816
Corresponding Measures:
Measure Title: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma
Measure Steward: University Hospitals Cleveland Medical Center
Brief Description of Measure: This measure estimates the proportion of emergency department (ED) visits that meet criteria for the ED being the appropriate level of care, among all ED visits for identifiable asthma in children and adolescents.
Developer Rationale: Asthma is one of the most common indications for emergency department (ED) visits by children. (1-3) AHRQ's Healthcare Cost and Utilization Project (HCUP) data from the Nationwide Emergency Department Sample (NEDS) found that in 2012, children between 1 and 17 years old had more than 1,895,000 ED visits for asthma with almost 10% resulting in hospitalization.

Evidence suggests that ED visits and hospitalizations in children with asthma vary systematically by how well-equipped that community is to provide primary care, and by the quality of primary care delivered. (4, 5) There is widespread literature illustrating that ED visits and hospitalizations are each undesirable utilization outcomes from poorly managed asthma. There is not a large literature that assesses whether or not pediatric ED visits were appropriate. (6 -10)

A body of literature has explored the value and feasibility of measuring the appropriateness of medical activities using data available in the medical record. (11-14) Early work in adults included assessment of hysterectomy, carotid endarterectomy and cardiac interventions. An independent research project brought the construct of appropriateness to children (15), while Kleinman and colleagues were the first to assess the appropriateness of specific pediatric procedures. (16, 17) A later study demonstrated the feasibility of medical record data for such an assessment. (18) DeAngelis pioneered studies of what constitutes a good reason to use the ED. (6) All of these studies used a definition of appropriateness that compared benefit to likely risk without specific consideration of costs. The need for more studies looking for overuse was recently reviewed. (19) RAND type Delphi panels are accepted around the world as a method for developing criteria to assess appropriateness. (20-22)

Research demonstrates that:

•ED visits are an important issue for child health insurers, including Medicaid, with clinical and financial consequences;

•An overcrowded primary care system contributes to ED use for non-emergent and even non-urgent conditions.

•Pediatric hospitalizations for asthma vary by primary care availability and quality

•ED visits are common for children with asthma, including those in Medicaid

•Assessment of appropriateness using information in the medical record is a well-established and validated method that has been successfully applied to children.

The literature suggests that a measure that assesses whether or not the ED is an appropriate level of care for a child with asthma at the time that they present has intrinsic value. Such a measure would:

•Characterize the process of care in a way that assesses whether a particular ED visit represents overuse

•Allow the outcomes of asthma care to be better characterized in a manner that describes performance and promotes targeted improvement. Inappropriate ED visits represent failures of primary care delivery, availability and/or access. Appropriate visits may represent a failure to control asthma. These have distinct and distinguishable meanings that contribute to the understanding of the quality of asthma care.

•Measuring the quality of asthma care requires assessment of multiple factors. This appropriateness measure helps plans, purchasers, and society to understand the implication of asthma ED visits as outcomes of asthma care. The implications herein is that understanding what is better or worse care requires looking at various factors and not simply a higher or lower appropriateness score. The understanding of this measure is enhanced by considering whether the rate of undesirable outcomes (ED visits and hospitalizations) is high or low and whether other measures of primary care availability and access or asthma quality suggest high levels of performance or not..

An abstract describing the proposed measure was peer-reviewed and subsequently presented to a national audience at AcademyHealth 2014 Annual Research Meeting in San Diego in the “Measuring the Safety, Quality, and Value” section. Feedback was positive regarding the methods, measures, ethics, and importance of this measure.

Research evidence supports the importance and need for our proposed measure that assesses whether the ED represents an appropriate level of care for children with asthma who are seen in the ED.

1.Kharbanda, A.B., et al., Variation in resource utilization across a national sample of pediatric emergency departments. J Pediatr, 2013. 163(1): p. 230-6.
2.Adams, J.G., Emergency department overuse: Perceptions and solutions. JAMA, 2013. 309(11): p. 1173-1174.
3.Institute, N.E.H., A Matter of Urgency: Reducing Emergency Department Overuse. Research Brief, 2010(March).
4.Perrin, J.M., et al., Variations in rates of hospitalization of children in three urban communities. N Engl J Med, 1989. 320(18): p. 1183-7.
5.Perrin, J.M., et al., Primary care involvement among hospitalized children. Arch Pediatr Adolesc Med, 1996. 150(5): p. 479-86.
6.DeAngelis, C., P. Fosarelli, and A.K. Duggan, Use of the emergency department by children enrolled in a primary care clinic. Pediatr Emerg Care, 1985. 1(2): p.61-5.
7.Berns, S.D., et al., Appropriate use of a pediatric emergency department: is the pediatrician called before the visit? Pediatr Emerg Care, 1994. 10(1): p. 13-7.
8.Rudowitz, R., A Look At CBO Projections For Medicaid and CHIP, in The Kaiser Commission on Medicaid and the Uninsured. 2014, The Henry J. Kaiser Family Foundation
9.Taubman, S.L., et al., Medicaid Increases Emergency-Department Use: Evidence from Oregon´s Health Insurance Experiment. Science, 2014. 343(6168): p. 263-268.
10.Smulowitz, P.B., et al., Increased Use of the Emergency Department After Health Care Reform in Massachusetts. Ann Emerg Med, 2014.
11.Brook, R.H., et al., A method for the detailed assessment of the appropriateness of medical technologies. Int J Technol Assess Health Care, 1986. 2(1): p. 53-63.
12.Park, R.E., et al., Physician ratings of appropriate indications for six medical and surgical procedures. Am J Public Health, 1986. 76(7): p. 766-72.
13.Fitch, K., et al., The RAND/UCLA Appropriateness Method User´s Manual. 2001 RAND.
14.Kosecoff, J., et al., The appropriateness of using a medical procedure. Is information in the medical record valid? Med Care, 1987. 25(3): p. 196-201.
15.Kemper, K.J., Medically inappropriate hospital use in a pediatric population. N Engl J Med, 1988. 318(16): p. 1033-7.
16.Kleinman, L.C., et al., The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States. Jama, 1994. 271(16): p. 1250-5.
17.Kleinman, L.C., E.A. Boyd, and J.C. Heritage, Adherence to prescribed explicit criteria during utilization review. An analysis of communications between attending and reviewing physicians. Jama, 1997. 278(6): p. 497-501.
18.Keyhani, S., et al., Electronic health record components and the quality of care. Med Care, 2008. 46(12): p. 1267-72.
19.Keyhani, S. and A.L. Siu, The underuse of overuse research. Health Serv Res, 2008. 43(6): p. 1923-30.

20.Bernstein, S.J., et al., The appropriateness of hysterectomy. A comparison of care in seven health plans. Health Maintenance Organization Quality of Care Consortium. Jama, 1993. 269(18): p. 2398-402.
21.Taylor, A.J., et al., ACCF/SCCT/ACR/AHA/ASE/ASNC/NASCI/SCAI/SCMR 2010 Appropriate Use Criteria for Cardiac Computed Tomography. A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, the Society of Cardiovascular Computed Tomography, the American College of Radiology, the American Heart Association, the American Society of Echocardiography, the American Society of Nuclear Cardiology, the North American Society for Cardiovascular Imaging, the Society for Cardiovascular Angiography and Interventions, and the Society for Cardiovascular Magnetic Resonance. J Cardiovasc Comput Tomogr, 2010. 4(6): p. 407.e1-33.
22.Basger, B.J., T.F. Chen, and R.J. Moles, Validation of prescribing appropriateness criteria for older Australians using the RAND/UCLA appropriateness method. BMJ Open, 2012. 2(5).

Numerator Statement: The numerator is the number of eligible asthma ED visits in the random sample that also satisfy at least one of the explicit criteria to indicate that the ED is an appropriate level of care. Distinct numerators are reported for children ages 2-5, 6-11, 12-18, and optionally, 19 - 21.
Denominator Statement: The denominator is a random sample of the patients in each age stratum who have visited the emergency department for asthma (as a first or second diagnosis) and meet the specified criteria for having identifiable asthma (defined in s2b).

Separate numerators and denominators are reported for children age 2-5, 6-11, 12-18, and, optionally, 19-21 years. An overall rate across strata is not reported.
Denominator Exclusions: ED visits that are already in the sample OR Children that fall outside of specified age range of 2-21 OR who do not meet time enrollment criteria OR do not meet identifiable asthma prior to the ED visit, OR children with concurrent or pre-existing COPD, Cystic Fibrosis or Emphysema.

At the discretion of the accountability entity, the denominator may be restricted to children 2-18.

Measure Type: Outcome
Data Source: Claims (Only), EHRs Hybrid, Paper Records
Level of Analysis: Facility, Health Plan

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary

This measure was previously reviewed by the Pulmonary Standing Committee (March 2016) as a process measure. Based on feedback from that Committee, the developer has revised and resubmitted the measure as an outcome measure. In the measure's evidence and testing forms, content submitted previously is in black; new information is in blue.

- The developer provides a diagram of the relationship between processes of care and outcomes.
- The developer states "Low levels of appropriateness suggest fewer breakthrough episodes of asthma and hence better quality of asthma care for those who receive it. If the rate of asthma ED visits is high and the rate of appropriateness is low this suggests both high quality care for those receive asthma care and insufficient access/availability of such care. High levels of appropriateness suggest both efficient resource use of the emergency department and that ED visits are a proxy for clinical outcomes since many of the visits represent

breakthrough asthma. High levels of appropriateness combined with a low rate of ED asthma use suggests both efficient use of resources and good asthma outcomes."

- The developer added citations for clinical practice guidelines from the National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (2007): "As a general rule, patients with well-controlled asthma should have: ... no emergency department visits; no hospital stays ...". Grade C = Nonrandomized trials and observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.
- The developer's criteria for appropriateness are listed as "explicit criteria" in the numerator details:
    - **Disposition of the ED visit was admission to the hospital, OR**
    - **Documented physical findings consistent with respiratory distress, including any of the following:** Labored breathing (including moderate or severe increased work of breathing); OR Retractions, grunting, and/or evidence of accessory muscle use; OR Markedly decreased breath sounds; **OR**
    - **Recorded oxygen saturation below 90%; OR**
    - **An arterial blood gas (ABG) was obtained in the emergency department; OR**
    - **The child had a consultation with a pulmonologist or asthma specialist that was ordered and provided in the ED; OR**
    - **There is clear documentation that prior to arrival in the ED any of the following occurred:** The child was referred to the ED after evaluation by the PCP or other clinician. The evaluation may include an in person visit or auscultation including via telephone OR The child received two or more doses of inhaled rescue medications without sufficient clinical improvement. Documentation of parent report meets the criterion. OR The child was assessed with an objective instrument such as a peak flow meter and was found to be in a pre-defined "red zone" of peak flow measurement as part of an asthma action or similar plan. Documentation requires ALL of the following: a written asthma action plan exists AND defines a "red zone" for which urgent assessment by a clinician is indicated AND an objective assessment was made and its result was in the pre-defined red zone. Documentation of parent report meets the criterion.
- NQF provides specific guidance on evaluating appropriate use measures, as follows:
    - "If there is no empiric evidence, skip Box 10 and go to Box 11. The Committee should agree that the AUC method is a systematic assessment of expert opinion that the benefits of what is being measured outweigh the potential harms (Box 11). If the Committee agrees that it is acceptable (or beneficial) to hold providers accountable for the performance in the absence of empiric evidence (Box 12), then rate as "insufficient evidence with exception."

Questions for the Committee:

- *Does the rationale provided by the developer support a relationship between appropriateness of ED visits and health outcomes?*
- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Guidance from the Evidence Algorithm: Outcome measure (Box 1) → Relationship between health outcome and provider action (Box 2) → PASS

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

Previous review: Submitted as process measure, did not pass Evidence - H-0; M-2; L-9; I-9

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

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

- The developer states, "asthma is one of the most common indications for emergency department (ED) visits by children". The developer reports on AHRQ data from 2012 indicating "children between 1 and 17 years old had more than 1,895,000 ED visits for asthma with almost 10% resulting in hospitalization." Further, they note that "evidence suggests that ED visits and hospitalizations in children with asthma vary systematically by how well-equipped that community is to provide primary care, and by the quality of primary care delivered. There is widespread literature illustrating that ED visits and hospitalizations are each undesirable utilization outcomes from poorly managed asthma."
- The developer reports the following results (from testing data) identified statistically significant differences between groups at specified levels, e.g., age groups, among racial/ethnic groups, and within age group among racial/ethnic groups:
    - 181 of 335 (54.3%) ED visits were deemed appropriate for children 2 to 5 years
    - 209 of 447 (43.8%) ED visits were appropriate for children 6 to 11 years
    - 165 of 341 (48.4%) visits were appropriate for children 12 to 18 years

Disparities

- The developer states, "Pediatric asthma is more prevalent in minority populations. Lifetime prevalence rates of asthma in Hispanic and African American children are 12.4% and 15.8% respectively."
- Based on its chart audits, the developer reports performance on the measure varies by race/ethnicity and that a Chi-square analysis confirms the differences are statistically significant. For example, Hispanic children had higher rates of questionable use of the ED (55.9% of visits) when compared to non-Hispanic children (47.8%), p=0.002. African American children "showed a trend" toward more questionable use compare to all other children (53.6% vs. 48.7%, p=0.10).
- The developer reports performance on the measure varies by insurance status and that a Chi-square analysis confirms the differences are statistically significant. The appropriate use rates were: Medicaid patients (46.3%); private (59%); uninsured patients (38.6%); other forms of insurance (military and worker's comp) (55.0%), p=0.005.

Question for the Committee:

- *Does the Committee believe there is a gap in care that warrants a national performance measure?*

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

Committee pre-evaluation comments

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

- I am concerned on multiple levels that this measure doesn't really look at appropriateness (appropriate use versus overuse and underuse) but rather focuses on overuse. The denominator is defined by patients who make it to the ER, but what about those that are not appropriately referred? The conceptual model also is based on ER use and appropriateness, but is not captured by the measure as proposed. Moreover, the measure assumes physician referral to the ER is defacto appropriate--I don't see quality evidence supporting this supposition. Other criteria for appropriateness beg quibbling: consultation, ABG (maybe a standing order), decreased breath sounds (reliable?), etc. Furthermore, the use of the second level diagnosis concerns me as it may capture ER visits with nothing to do with asthma. The causal pathway here seems fraught with confounders. And the measure as proposed inadequate to determine appropriate use. I do not support an exception.
- Substantial evidence is provided that demonstrates the improvement in overall health for these chronically ill patients if appropriate asthma visits are reduced as they are usually patients in crisis. As an outcome measure we need to assess whether it is actionable for the unit of measurement. The developers say it can be used by hospitals and insurers/health plans. Further evidence should be provided on how this is actionable at the hospital level as the primary interventions would be associated with PCPs, specialists and urgent care facilities. For a health plan, this data will be sampled from multiple EDs yet this is not part of the sampling algorithm. This can be actionable by the insurer as they control the network.
- I would agree there is a gap in care, and certainly evidence for disparities driving outcomes.
- Most of the rates reported were around the 40-50% range which indicates substantial room for improvement.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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

Data source(s): Administrative claims, Electronic Clinical Data: Electronic Health Record, Paper Medical Records

Specifications:

- The level of analysis is facility and health plan; the care setting is ED, hospital.
- Interpretation of score: Better quality = Higher score
- The numerator for this measure is: *The numerator is the number of eligible asthma ED visits in the random sample that also satisfy at least one of the explicit criteria to indicate that the ED is an appropriate level of care. Distinct numerators are reported for children ages 2-5, 6-11, 12-18, and optionally, 19 - 21.*
- ***NQF Note:***
  - The listed "explicit criteria" in the specifications (numerator details) are:
    - **Disposition of the ED visit was admission to the hospital, OR**
    - **Documented physical findings consistent with respiratory distress, including any of the following:** Labored breathing (including moderate or severe increased work of breathing); OR Retractions, grunting, and/or evidence of accessory muscle use; OR Markedly decreased breath sounds; **OR**
    - **Recorded oxygen saturation below 90%; OR**
    - **An arterial blood gas (ABG) was obtained in the emergency department; OR**
    - **The child had a consultation with a pulmonologist or asthma specialist that was ordered and provided in the ED; OR**
    - **There is clear documentation that prior to arrival in the ED any of the following occurred:** The child was referred to the ED after evaluation by the PCP or other clinician. The evaluation may include an in person visit or auscultation including via telephone OR The child received two or more doses of inhaled rescue medications without sufficient clinical improvement. Documentation of parent report meets the criterion. OR The child was assessed with an objective instrument such as a peak flow meter and was found to be in a pre-defined "red zone" of peak flow measurement as part of an asthma action or similar plan. Documentation requires ALL of the following: a written asthma action plan exists AND defines a "red zone" for which urgent assessment by a clinician is indicated AND an objective assessment was made and its result was in the pre-defined red zone. Documentation of parent report meets the criterion.
- The denominator for this measure is: *The denominator is a random sample of the patients in each age stratum who have visited the emergency department for asthma (as a first or second diagnosis) and meet the specified criteria for having identifiable asthma (defined in s2b).*
  Separate numerators and denominators are reported for children age 2-5, 6-11, 12-18, and, optionally, 19-21 years. ***An overall rate across strata is not reported*** [our emphasis].
- The exclusions for the measure are: *ED visits that are already in the sample OR Children that fall outside of specified age range of 2-21 OR who do not meet time enrollment criteria OR do not meet identifiable asthma prior to the ED visit, OR children with concurrent or pre-existing COPD, Cystic Fibrosis or Emphysema.*
  At the discretion of the accountability entity, the denominator may be restricted to children 2-18.
- The ICD-9 and ICD-10 codes are in an appendix.
- The calculation algorithm is stated in S.14.
- Sampling is allowed. At least 500 children per strata should be included in the samples.
- One data source is pharmacy claims, but the developer acknowledges that availability will vary.
- The measure requires stratification by three age groups, as just noted.

Questions for the Committee:

- *Are all the data elements clearly defined? Are the appropriate codes included in the ICD-9 to ICD-10 conversion?*
- *Is the calculation algorithm clear?*
- *Is the potential variability in access to/inclusion of pharmacy data a concern?*
- *Is it a concern that the measure does not report an overall rate?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

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

SUMMARY OF TESTING

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

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

Method(s) of reliability testing

- The developer relied on pre-existing data element-level validity testing in the literature to identify children who are being managed for identifiable asthma (denominator), which is permitted by NQF testing guidance.
- For the numerator, the developer did not conduct empirical reliability testing. Instead, it relied on empirical validity testing at the data element level (chart abstraction compare to an authoritative source).
- Per NQF guidance, separate reliability testing is not required if data element-level validity testing is performed.

Results of reliability testing

- Not applicable; see discussion on validity testing at the data element level

Question for the Committee:

o *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm Not Applicable (rating from data-element validity will apply; highest eligible rating is MODERATE)

Precise specifications (Box 1) → Empirical reliability testing conducted using statistical tests (Box 2) → Empirical validity testing of patient-level data conducted (Box 3) → Insufficient

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

Rationale: See rationale for validity.

2b. Validity

2b1. Validity: Specifications

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

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

o *Are the specifications consistent with the evidence?*

2b2. Validity testing

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

SUMMARY OF TESTING

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

Method of validity testing of the measure score:

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

Validity testing method:

- The developer reports data element level validity testing. The developer relies, as is permitted by NQF guidance, on other sources for denominator data element validity and conducted empirical testing on the numerator data elements in one facility.

For the denominator:

- The developer relies on literature to support its conclusion of the validity of administrative data elements to identify children who are being managed with identifiable asthma. Per NQF policy:
  - Prior evidence of validity of data elements can be used, including published data, provided it includes the same data elements; uses the same data type; and is conducted on an appropriate sample (i.e., representative, adequate numbers, etc.)
  - The developer attests that the data elements match those assessed in the literature.
- The developer also cites score-level validity testing of two previously-endorsed asthma measures as evidence of data-element level validity. However, this does not meet NQF's requirements for demonstration of score-level validity.
- The developer used NY State Medicaid Managed Care claims data for its analyses.

For the numerator:

- For validity testing at the data element level, three reviewers each looked at 10 charts from one facility, assessing the presence of 6 constructs and an overall visit-level assessment of appropriateness. The developer states it conducted testing at the beginning of data collection and again at the conclusion of data collection.
- The following six numerator appropriateness criteria were tested:
  - Retractions
  - Accessory Muscle Use
  - Markedly diminished BS
  - Hospitalized from ED
  - O2 sat < 90%
  - Referred by PCC

Validity testing results:

Denominator

- For the results of the literature review:
  - The developer attests that the data elements match those assessed in the literature.
  - The sensitivities, specificities, PPVs, kappas, etc., were generally strong, including:
    - Wilchesky et al., asthma diagnosis in in-patient setting: Sp= 96.76 (95%CI 96.5, 97.0)
    - Folwes et al., asthma diagnosis in ambulatory: Sensitivity and specificity was 0.82 and 0.99, respectively. Sensitivity of .82 using claims was higher than sensitivity using self-report at 0.64
    - Wilchesky, et al., asthma diagnosis in clinic/outpatient setting: Asthma claims were highly specific, Sp= 96.76 (95%CI 96.5, 97.0).
- The developer also provides information from various articles related to the use of administrative data for identification of asthma and use of claims data for performance measurement.
  - Age: According to CMS MMIS data requirements, "States are required to submit validated claims data including age or date of birth with a tolerance of 0.1%".

- Asthma diagnosis: In an in-patient/ED setting, "Asthma claims were highly specific, Sp= 96.76 (95%CI 96.5, 97.0)." (Wilchesky, et al)
- In an ambulatory setting, "Sensitivity and specificity was 0.82 and 0.99, respectively. Sensitivity of .82 using claims was higher than sensitivity using self-report at .64" (Folwes, et al)
- In a clinic/outpatient setting, "Asthma claims were highly specific, Sp= 96.76 (95%CI 96.5, 97.0)." (Wilchesky, et al)
- Exclusions: for diagnoses of COPD, cystic fibrosis, emphysema, "Claims had a PPV of 91.9, and a negative predictive value of 92.6, with k of 0.65 (substantial agreement) compared to chart review for chronic pulmonary disease. ICD 10 performed similarly in this study" (Quan et al)

Numerator

- For the six constructs, three reviewers each reviewed 10 charts early in training and at the end of the data collection period. The developer reports this resulted in 180 comparisons with the trainer (6 clinical constructs * 3 * 10 = 180). The table also provides an "all six combined" kappa, but it is unclear why this is lower than the overall assessment and its implications for reliability. NQF staff has requested additional information from the developer.
- For the assessment of overall appropriateness (i.e., a separate assessment and the numerator), the kappas were 0.77, increasing to 0.87 after training.

The developer states, per the Landis and Koch classification, a kappa value of 0.87 indicates almost perfect agreement. **Construct**	**Agreement**	
	Initial Kappa	**Final Kappa**
1. Retractions	0.67	0.87
2. Accessory Muscle Use	0.44	0.89
3. Markedly diminished BS	0.71	0.78
4. Hospitalized from ED	1.0	1.0
5. O2 sat < 90%	0.79	NA*
6. Referred by PCC	1.0	NA*
All six combined	0.76	0.68
Overall: Appropriateness	0.77	0.87
* NA is because there was no variability in the charts reviewed. There was no disagreement in any of the assessments.		

- The explicit criteria for the numerator specifications include several elements for which results are not reported—e.g., "arterial blood gas (ABG) in the emergency department".
  - In an email to NQF staff, the developer stated it tested all elements, but "only reported findings from the criteria that were found to be pertinent within our tested institution. Because these appropriateness criteria are written for implementation nationally, and there is variation among hospitals protocols/procedures done in the ED, we included all the appropriateness criteria."
  - The developer further stated that at the institution where the measure was tested, "ABG ordered/obtained were not reported because ABGs are not ordered in the ED [at this institution], and rely on the PulseOx O2 Saturation level instead. But the criteria included in the specifications are written for national implementation."

Questions for the Committee:

- *Is the test sample size adequate to generalize for widespread implementation?*
- *Should all numerator appropriateness criteria be tested for data-element level validity?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer provides the following:

- There are no numerator exclusions.
- Denominator exclusions include: Children with concurrent or pre-existing Chronic Obstructive Pulmonary Disease (COPD) diagnosis, cystic fibrosis diagnosis, or emphysema diagnosis.
- The developer reports <=2.5% potentially eligible children were excluded by these clinical diagnoses.
- The developer reports that exclusions are clinical and represent construct validity rather than statistical considerations.
- The measure also excludes children who have not been consecutively enrolled in the reporting plan for at least two months prior to the index reporting month, as well as the index reporting month itself, but they note that 20% more children are included than would be if they had a 12-month enrollment requirement.

Questions for the Committee:

- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Are the exclusions/exceptions of sufficient frequency and variation across providers to be needed (and outweigh the data collection burden)?*

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

Conceptual rationale for SDS factors included? ☐ **Yes** ☒ **No**

SDS factors included in risk model? ☒ **Yes** ☐ **No**

Risk adjustment summary

- An overall rate is not reported; the measure is reported by age stratifications.
- This measure is adjusted for age group only (ages 2-5, 6-11, 12-18, and optionally, 19-21). The developer has specified stratified analysis for risk-adjustment rather than using a statistical risk modeling approach.
- The developer relies on the NIH NHLBI NAEPP guideline assertion that the goals for asthma severity, control, and responsiveness are identical for all levels of baseline asthma severity as the rationale not to risk-adjust for severity.
- The developer states that additional stratifications are optional, but notes that these were not included as risk adjustment factors due to lack of "*clear biological evidence that ED visits should be more likely in any of the sub categories*".

Conceptual analysis of the need for SDS adjustment:

- Although the developer noted that its funders asked them to consider SDS factors, and it in fact did find patient-level differences for at least some, the developer did not discuss the conceptual rationale of why or how SDS factors (e.g., race/ethnicity, poverty level in the caregivers county of residence, rurality/urbanicity on the caregiver's county of residence, insurance type and plan type) might be associated with appropriateness of ED visits for asthma.

Empirical analysis of SDS factors:

- The developer reported statistically significant differences in appropriateness of ED visits at the patient level by race and ethnicity, as well as for sex and insurance status subgroups. It did not, however, compare rates of appropriateness with and without inclusion of these variables as a risk-adjustment approach, as requested by NQF.

Questions for the Committee:

- *A justification for no risk adjustment for SDS factors is provided. Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors? Is there evidence that contradicts the developer's rationale?*
- *Do you agree with the developer's decision not to adjust for severity?*

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

- In its testing data, the developer tested the numerator at 1 facility, so cannot demonstrate meaningful differences at the proposed facility level of analysis.
- The developer also analyzed claims data from the New York State Medicaid Managed Care data (including claims from all MCOs that are contracted for Medicaid care), so could not analyze differences among (for example) state Medicaid programs. The developer does not examine differences among MCOs within the data plan.
    - The developer states its analyses found meaningful differences by age groups and statistically significant differences by race/ethnicity and insurance status, which the developer states means "the measure distinguishes signal from noise". The developer posits that this demonstrates the measure detects meaningful differences. NQF's requirement is that testing demonstrate differences among measured entities.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

n/a

2b7. Missing Data

- The developer does not account for missing data. It cites literature that chart review is an accurate method of identifying the level of appropriateness of a clinical service. Failure to document is a "quality deficit" that the developer does not consider as missing data.
- Use of pharmacy data is on an "if available" basis to identify children with asthma for the denominator; the developer notes any results reported without should be marked as such. The developer reports use of pharmaceutical data expanded the pool by approximately 10,000 children (from 180,000 to 190,000—5.5%). The developer states it "found no evidence this was a threat to validity," but does not provide analyses that the scores with the pharmacy data did not differ from the scores when pharmacy data were excluded. The developer does not have direct access to the data to provide additional analyses at this time.

Question for the Committee:

- *Is the variable use of pharmacy data a threat to validity?*

Guidance from the Validity Algorithm: Specifications consistent with evidence (Box 1) → Potential threats to validity assessed (Box 2)→ Insufficient

The highest possible rating is INSUFFICIENT.

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

RATIONALE: The numerator was only tested at one facility with 10 charts; the measure does not demonstrate meaningful differences. All numerator details/data elements were not tested (e.g., ABG was not available at the institution at which testing was conducted, but its reliability/kappa should be assessed at an institution for which this is policy/practice). Additionally, there is insufficient information for the Committee to discuss whether SDS factors should

or should not be included (i.e., analyses on scoring with and without factors included). Finally, there are insufficient data on the effect on the score of missing pharmacy data.

Committee pre-evaluation comments

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

- I am worried that there are substantial challenges in collecting this data consistently, and potential for elements such as pharmacy data systematically influencing outcomes. The outcome measure itself is really not useful as a quality measure without knowing utilization overall.
- This is being presented as an Outcome Measure. Substantial evidence is provided concerning the denominator definition and the codes seem appropriate. The numerator was determined by the RAND/UCLA Delphi method with a panel of experts. If appropriately used this is considered a best approach to applying the Delphi method. Even though its primarily for designing survey questions it is still appropriate for determining the list of indicators for inappropriate use. A clinician should address the appropriateness of the panel measures chosen
- The developers say this will be done like a HEDIS hybrid measure where a stratified sample will be used to generate members of the denominator and the claims and chart review will determine the numerator. Will discuss the validity of the numerator method in that section but it was clear that even in this one hospital pilot, some of the numerator indicators are not collected. The developers say that is okay as the collected indicators are enough. Hard to judge without multiple institutions.
- The algorithm is clear for sampling the denominator (except for exact sample sizes) and it is the “at least one indication” for the numerator.
- The measure is reported stratified by age based on data provided that show age variability on this rate exists. This is appropriate. There is no data to support consistent implementation though the steps to determine the rate are well specified.
- There was no specific reliability testing done. However acceptable validity results on the denominator were provided. Validity testing of the numerator was done at the item level comparing chart to an authoritative source with mixed results.
- I guess according to NQF criteria it is sufficient, although I am very concerned that putting apples, oranges, and pineapples together gets you fruit cocktail...
- There was no specific reliability testing done. However acceptable validity results on the denominator were provided. Validity testing of the numerator was done at the item level comparing chart to an authoritative source with mixed results. Only 30 charts were reviewed at one institution which is insufficient to generalize the results. Some items could not be assessed as they were not recorded in the charts at this institution which makes item level testing insufficient. Kappas were okay for some items but not others. They improved over time but does that imply that the measure should be “practiced” the first year and not used until year 2?The comparison of chart to authoritative source could be assessing: Accuracy of chart data or The level of documentation detail provided by the clinician or The Quality of the data abstractors. No assessment was done comparing the charts to the claims data to assess whether there may be inaccuracies in the claims data. This can be considered a measure of quality but not necessarily at the hospital or health plan level. Rather it is an outcome that can be used for interventions at the provider level. It should be noted that they surveyed nine other institutions to see if they think any of the chart abstraction or item specifications would be problematic and results were consistently positive for successful data collection.
- I am concerned about the pharmacy factor, and also varying availability of data within the specification such as ABG. The testing is not very comforting...
- "Exclusion criteria seems acceptable and should not be burdensome as these can be identified from the claims data. The reasoning for no risk adjustment provided by the developers is sound if this measure is to be used for Quality Improvement as opposed to benchmarking (typically HEDIS measures are for both). QI interventions should be done based on the true population of the hospital or insurer. Interventions might be different in a poorer region vs one that is not or in an area with a large minority population. If benchmarking is also a goal then a risk adjusted rate should be calculated as well given the differences in rates by subgroups that was presented. Meaningful differences across institutions could not be adequately assessed based on the evidence provided. This is critical if the rates will be used for benchmarking but less so if for local quality improvement initiatives. Missing pharmacy data was not adequately addressed as no empirical data showing lack of impact was provided.

Criterion 3. Feasibility

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

- The developer reports some data elements are in defined fields in electronic sources.
- The developer reports there are no fees.
- The measure requires chart abstraction, and testing found that training improves data element level validity/reliability. The developer notes that chart review is a reliable and accepted method of measuring appropriate use. No information is provided on the minimum number of charts that should be assessed.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Is the data collection strategy ready to be put into operational use?*
- *What training is available, how is it accessed, and what are the costs associated with training a clinician to identify an event? Is there additional burden for that physician?*
- *Can trained nurses or nurse practitioners review records?*

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

Committee pre-evaluation comments

Criteria 3: Feasibility

- I think this is feasible but very burdensome.
- Given the similarity to HEDIS measures, the data collection should be feasible given the needed training. Trained nurse or nurse practitioners should be able to do the chart review and there should not be additional burden for the physician unless the hospital adds fields to the EMR to improve validity of the rate.

Criterion 4: Usability and Use

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details

- The developer plans to assist in the implementation of this measure following NQF endorsement. The developer notes multiple stakeholders are interested in using the measure. It has been approved for inclusion in the National Quality Measures Clearinghouse.

Improvement results

- This measure is not in use and therefore, has no improvement results.

Potential harms

- The developer reports no unintended consequences were observed during testing.
- The developer notes that the measure has a lower risk for gaming than some other measures because both low and high results can demonstrate different areas for improvement.

Vetting of the measure

- N/A

Feedback:

- No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal program.

***Questions for the Committee**:*

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

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

Committee pre-evaluation comments

Criteria 4: Usability and Use

- I am not sold on the link of this measure to quality as noted many times above.
- As designed, this measure can be used to assess improvements in the care these asthma patients.

Criterion 5: Related and Competing Measures

Related or competing measures

The developer did not include information on any of the related or competing measures. However, NQF staff identified the following measures that may be related and/or competing.

- 0047: Asthma: Pharmacologic Therapy for Persistent Asthma
- 0728: Asthma Admission Rate (PDI 14)
- 1800: Asthma Medication Ratio
- 2414: Pediatric Lower Respiratory Infection Readmission Measure
- 3189: Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: Visits per 100 Child-years (*submitted by the same developer for review in this project)*

Harmonization

No information available.

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

RATIONALE IF NOT ELIGIBLE:

The measure has not been vetted.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: 12/9/2016

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

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

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☒ Health outcome: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

☐Patient-reported outcome (PRO): Click here to name the PRO

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

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

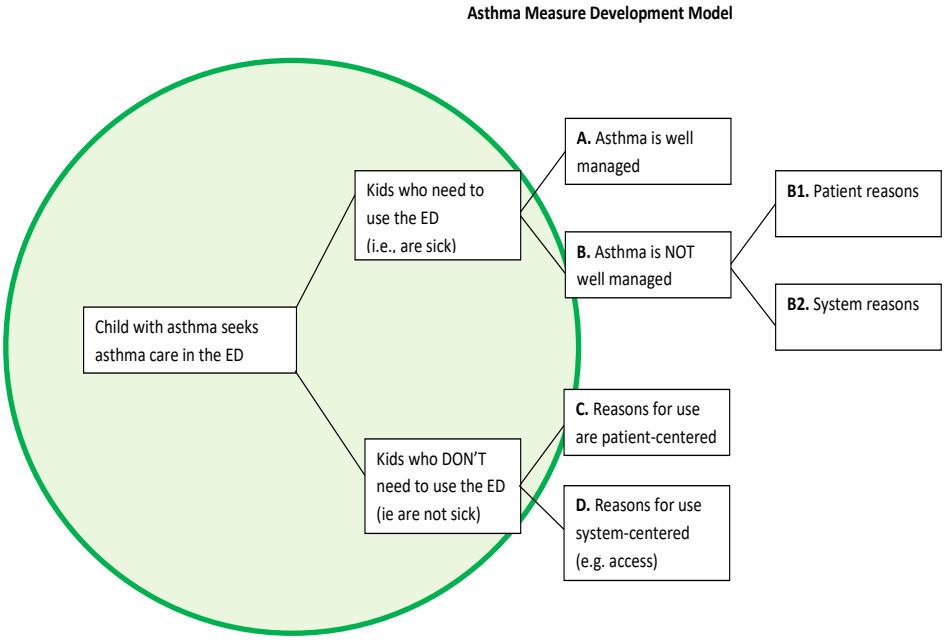
☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.



The green circle highlights the aspects of the conceptual model incorporated into this measure.
Underlying this model is a simple framework:

1. Accessible high quality primary care reduces the need for ED visits by decreasing the number of children who have acute breakthrough episodes requiring the ED. (NHLBI Guideline idenitifies factors with Evidence Levels A, B, and C).
2. Accessible high quality primary care reduces the need for ED visits by decreasing the number of children who come to the ED for asthma care better performed in the office setting.
3. Some children in the ED need to be there. Of those, some episodes were potentially preventable and others were not.
4. Our focus groups highlighted that some parents are comforted by the setting of the ED when they are caring for what they perceive as a vulnerable child. Parent perspectives do not adhere to system perspectives regarding a more strict hierarchy of what care belongs where.

Low levels of appropriateness suggest fewer breakthrough episodes of asthma and hence better quality of asthma care for those who receive it. If the rate of asthma ED visits is high and the rate of appropriateness is low this suggests both high quality care for those receive asthma care and insufficient access/availability of such care.

High levels of appropriateness suggest both efficient resource use of the emergency department and that ED visits are a proxy for clinical outcomes since many of the visits represent breakthrough asthma. High levels of appropriateness combined with a low rate of ED asthma use suggests both efficient use of resources and good asthma outcomes.

A seminal article that began to link primary care services to outcomes at a community or population level and that supports our interpretation appeared in the New England Journal of Medicine in 1989. (N Engl J Med. 1989 May 4;320(18):1183-7. Variations in rates of hospitalization of children in three urban communities.
Perrin JM1, Homer CJ, Berwick DM, Woolf AD, Freeman JL, Wennberg JE.).

Low levels of appropriateness suggest better asthma outcomes with less efficiency of primary care.
High levels of appropriateness suggest more efficient primary care and worse asthma outcomes.

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

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

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

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

x Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• National Heart, Lung, and Blood Institute, National Institutes of Health (NHLBI/NIH) Asthma Guideline 2007 • www.nhlbi.nih.gov/guidelines/asthma (NAEPP Guideline)
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline,	The impact of asthma management and access to care on ED outcomes and clinical control are described well in the NHLBI NAEPP guideline. The Clinical guideline acknowledges ED visits as failures of control.

summarize the conclusions from the SR.	Class C Evidence is suggested when suggesting the relationship between specific periodicity of ambulatory visits with asthma outcomes but higher levls of evidence including Class A with the relationship between treatment and outcomes. Quick Reference: Asthma control focuses on two domains: 1)reducing impairment--- the frequency and intensity of symptoms… and 2) reducing risk – the likelihood of future asthma attacks… [later described as "prevent exacerbations] NHLBI Guideline: *As a general rule, patients with well-controlled asthma should have:* • *Few, if any, asthma symptoms.* • *Few, if any, awakenings during the night caused by asthma symptoms.* • *No need to take time off from school or work due to asthma.* • *Few or no limits on full participation in physical activities.* • *No emergency department visits.* • *No hospital stays.* • *Few or no side effects from asthma medicines.*
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The National Asthma Education and Prevention Program (NAEPP) guidelines are the prevailing clinical recommendation for children with asthma. The Expert Panel Reports presenting clinical practice duielines for the diagnosis and management of asthma have organized recommendations for asthma care around four components considered essential to effective asthma management: - Measures of assessment and monitoring, obstained by objective tests, physical examination, patient history and patient report, to diagnose and assess the characteristics and severity of asthma and to monitor whether asthma control is achieved and maintained. - Education for partnership in asthma care - Control of environmental factors and comorbid conditions that affect asthma - Pharmacologic therapy This section of the report updates information on each of these four components based on the Expert Panel's review of the scientific literature. The sections that follow present specific clinical recommendations for managing asthma long term and for managing exacerbations that incorporate the four compoenents.

Provide all other grades and definitions from the evidence grading system	Methodology for report: Overall Methods Used To Develop This Report **Background** In June 2004, the Science Base Committee of the NAEPP recommended to the NAEPP CC that its clinical practice guidelines for the diagnosis and management of asthma be updated. In September, under the leadership of Dr. Barbara Alving, M.D. (Chair of the NAEPP CC, and Acting Director of the NHLBI), a panel of experts was selected to update the clinical practice guidelines by using a systematic review of the scientific evidence for the treatment of asthma and consideration of literature on implementing the guidelines. In October 2004, the Expert Panel assembled for its first meeting. Using EPR-2 1997 and EPR-Update 2002 as the framework, the Expert Panel organized the literature searches and subsequent report around the four essential components of asthma care, namely: (1) assessment and monitoring, (2) patient education, (3) control of factors contributing to asthma severity, and (4) pharmacologic treatment. Subtopics were developed for each of these four broad categories. The steps used to develop this report include: (1) completing a comprehensive search of the literature; (2) conducting an indepth review of relevant abstracts and articles; (3) preparing evidence tables to assess the weight of current evidence with respect to past recommendations and new and unresolved issues; (4) conducting thoughtful discussion and interpretation of findings; (5) ranking strength of evidence underlying the current recommendations that are made; (6) updating text, tables, figures, and references of the existing guidelines with new findings from the evidence review; (7) circulating a draft of the updated guidelines through several layers of external review, as well as posting it on the NHLBI website for review and comment by the public and the NAEPP CC, and (8) preparing a final-report based on consideration of comments raised in the review cycle. **Preparation Of Evidence Tables** Evidence tables were prepared for selected topics. It was not feasible to generate evidence tables for every topic in the guidelines. Furthermore, many topics did not have a sufficient body of evidence or a sufficient number of high-quality studies to warrant the preparation of a table. The Panel decided to prepare evidence tables on those topics for which an evidence table would be particularly useful to assess the weight of the evidence-e.g., topics with numerous articles, conflicting evidence, or which addressed questions raised frequently by clinicians. Summary findings on topics without evidence tables, however, also are included in the updated guidelines text.

Evidence tables were prepared with the assistance of a methodologist who served as a consultant to the Expert Panel. Within their respective committees, Expert Panel members selected the topics and articles for evidence tables. The evidence tables included all articles that received a "yes" vote from both the primary and secondary reviewer during the systematic literature review process. The methodologist abstracted the articles to the tables, using a template developed by the Expert Panel. The Expert Panel subsequently reviewed and approved the final evidence tables. A total of 20 tables, comprising 316 articles are included in the current update (see figure 1-1). Evidence tables are posted on the NHLBI Web site.

Ranking The Evidence

The Expert Panel agreed to specify the level of evidence used to justify the recommendations being made. Panel members only included ranking of evidence for recommendations they made based on the scientific literature in the current evidence review. They did not assign evidence rankings to recommendations pulled through from the EPR-2 1997 on topics that are still important to the diagnosis and management of asthma but for which there was little new published literature. These "pull through" recommendations are designated by EPR-2 1997 in parentheses following the first mention of the recommendation. For recommendations that have been either revised or further substantiated on the basis of the evidence review conducted for the EPR-3: Full Report 2007, the level of evidence is indicated in the text in parentheses following first mention of the recommendation. The system used to describe the level of evidence is as follows (Jadad et al. 2000):

- **Evidence Category A:** Randomized controlled trials (RCTs), rich body of data. Evidence is from end points of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants.

- **Evidence Category B:** RCTs, limited body of data. Evidence is from end points of intervention studies that include only a limited number of patients, post hoc or subgroup analysis of RCTs, or meta-analysis of RCTs. In general, category B pertains when few randomized trials exist; they are small in size, they were undertaken in a population that differs from the target population of the recommendation, or the results are somewhat inconsistent.

- **Evidence Category C:** Nonrandomized trials and observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.

- **Evidence Category D:** Panel consensus judgment. This category is used only in cases where the provision of some guidance was deemed valuable, but the clinical literature addressing the

subject was insufficient to justify placement in one of the other categories. The Panel consensus is based on clinical experience or knowledge that does not meet the criteria for categories A through C.

In addition to specifying the level of evidence supporting a recommendation, the Expert Panel agreed to indicate the strength of the recommendation. When a certain clinical practice "is recommended," this indicates a strong recommendation by the panel. When a certain clinical practice "should, or may, be considered," this indicates that the recommendation is less strong. This distinction is an effort to address nuances of using evidence ranking systems. For example, a recommendation for which clinical RCT data are not available (e.g., conducting a medical history for symptoms suggestive of asthma) may still be strongly supported by the Panel. Furthermore, the range of evidence that qualifies a definition of "B" or "C" is wide, and the Expert Panel considered this range and the potential implications of a recommendation as they decided how strongly the recommendation should be presented.

Panel Discussion

The first opportunity for discussion of findings occurred within the "topic teams." Teams then presented a summary of their findings during a conference call to all members of their respective committee. A full discussion ensued on each topic, and the committee arrived at a consensus position. Teams then presented their findings and the committee position to the full Expert Panel at an in-person meeting, thereby engaging all Panel members in critical analysis of the evidence and interpretation of the data.

A series of conference calls for each of the 10 committees as well as four in-person Expert Panel meetings (held in October 2004, April 2005, December 2005, and May 2006) were scheduled to facilitate discussion of findings and to dovetail with the three cycles of literature review that occurred over the 18-month period. Potential conflicts of interest were disclosed at the initial meeting.

Report Preparation

Development of the EPR-3: Full Report 2007 was an iterative process of interpreting the evidence, drafting summary statements, and reviewing comments from the various external reviews before completing the final report. In the summer and fall of 2005, the various topic teams, through conference calls and subsequent electronic mail, began drafting their assigned sections of the report. Members of the respective committees reviewed and revised team drafts, also by using conference calls and electronic mail. During the calls, votes were taken to ensure agreement with final conclusions and recommendations. During the December 2005 meeting, Panel members reviewed and discussed all committee drafts.

During the May 2006 meeting, the Panel conducted a thorough review and discussion of the report and reached consensus on the recommendations. For controversial topics, votes were taken to ensure that each individual's opinion was considered. In July, using

	conference calls and electronic mail, the Panel completed a draft of the EPR-3: Full Report 2007 for submission in July/August to a panel of expert consultants for their review and comments. In response to their comments, a revised draft of the EPR-3: Full Report 2007 was developed and circulated in November to the NAEPP Guidelines Implementation Panel (GIP) for their comment. This draft was also posted on the NHLBI Web site for public comment in February 2007. The Expert Panel considered 721 comments from 140 reviewers. Edits were made to the documents, as appropriate, before the full EPR-3: Full Report 2007 was finalized and published. The EPR-3: Full Report 2007 will be used to develop clinical practice guidelines and practice-based tools as well as educational materials for patients and the public. **References** 1. EPR. Expert panel report: guidelines for the diagnosis and management of asthma (EPR 1991). NIH Publication No. 91-3642. Bethesda, MD: U.S. Department of Health and Human Services; National Institutes of Health; National Heart, Lung, and Blood Institute; National Asthma Education and Prevention Program, 1991. 2. EPR-2. Expert panel report 2: guidelines for the diagnosis and management of asthma (EPR-2 1997). NIH Publication No. 97-4051. Bethesda, MD: U.S. Department of Health and Human Services; National Institutes of Health; National Heart, Lung, and Blood Institute; National Asthma Education and Prevention Program, 1997. 3. EPR-Update 2002. Expert panel report: guidelines for the diagnosis and management of asthma. Update on selected topics 2002 (EPR-Update 2002). NIH Publication No. 02 5074. Bethesda, MD: U.S. Department of Health and Human Services; National Institutes of Health; National Heart, Lung, and Blood Institute; National Asthma Education and Prevention Program, June 2003. 4. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, Stevens R. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. BMJ 2000;320(7234):537-40. 5. NHIS. National health interview survey (NHIS 2005). Hyattsville, MD: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention, 2005. Available at http://www.cdc.gov/nchs/about/major/nhis/reports_2005.htm. Link to the evidence tables themselves: http://www.nhlbi.nih.gov/health-pro/guidelines/current/asthma-guidelines/evidence-tables
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence:	**Systematic Evidence Review Overview**

• Quantity – how many studies? • Quality – what type of studies?	**Inclusion/Exclusion Criteria** The literature review was conducted in three cycles over an 18-month period (September 2004 to March 2006). Search strategies for the literature review initially were designed to cast a wide net but later were refined by using publication type limits and additional terms to produce results that more closely matched the framework of topics and subtopics selected by the Expert Panel. The searches included human studies with abstracts that were published in English in peer reviewed medical journals in the MEDLINE database. Two timeframes were used for the searches, dependent on topic: January 1, 2001, through March 15, 2006, for pharmacotherapy (medications), peak flow monitoring, and written action plans, because these topics were recently reviewed in the EPR-Update 2002; and January 1, 1997, through March 15, 2006, for all other topics, because these topics were last reviewed in the EPR-2 1997. **Search Strategies** Panel members identified, with input from a librarian, key text words for each of the four components of care. A separate search strategy was developed for each of the four components and various key subtopics when deemed appropriate. The key text words and Medical Subject Headings (MeSH) terms that were used to develop each search string are found in an appendix posted on the NHLBI Web site. **Literature Review Process** The systematic review covered a wide range of topics. Although the overarching framework for the review was based on the four essential components of asthma care, multiple subtopics were associated with each component. To organize a review of such an expanse, the Panel was divided into 10 committees, with about 4-7 reviewers in each (all reviewers were assigned to 2 or more committees). Within each committee, teams of two ("topic teams") were assigned as leads to cover specific topics. A system of independent review and vote by each of the two team reviewers was used at each step of the literature review process to identify studies to include in the guidelines update. The initial step in the literature review process was to screen titles from the searches for relevancy in updating content of the guidelines, followed by reviews of abstracts of the relevant titles to identify those studies meriting full-text review based on relevance to the guidelines and study quality. The combined number of titles screened from cycles 1, 2, and 3 was 15,444. The number of abstracts and articles reviewed for all three cycles was 4,747. Of these, 2,863 were voted to the abstract Keep list following the abstract-review step. A database of these abstracts is posted on the NHLBI Web site. Of these abstracts, 2,122 were advanced for full-text review, which resulted in 1,654 articles serving as a bibliography of references used to update the guidelines, available on the NHLBI Web site. Articles were selected

	from this bibliography for evidence tables and/or citation in the text. In addition, articles reporting new and particularly relevant findings and published after March 2006 were identified by Panel members during the writing period (March 2006-December 2006) and by comments received from the public review in February 2007.
Estimates of benefit and consistency across studies	In summary, the NAEPP "Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma-Full Report 2007" represents the NAEPP's ongoing effort to keep recommendations for clinical practice up to date and based upon a systematic review of the best available scientific evidence by a Panel of experts, as well as peer review and critique by the collective expertise of external research/science consultants, the NAEPP CC members, guidelines implementation specialists, and public comment. The relationship between guidelines and clinical research is a dynamic one, and the NAEPP recognizes that the task of keeping guidelines' recommendations up to date is an increasing challenge. In 1991, many recommendations were based on expert opinion because there were only limited randomized clinical trials in adults, and almost none in children, that adequately tested clinical interventions grounded in research findings about the disease process in asthma. The large gaps in the literature defined pressing clinical research questions that have now been vigorously addressed by the scientific community, as the size of the literature reviewed for the current report attests. The NAEPP is grateful to all of the Expert Panel members for meeting the challenge with tremendous dedication and to Dr. William Busse for his outstanding leadership. The NAEPP would particularly like to acknowledge the contributions of Dr. Gail Shapiro, who served on NAEPP Expert Panels from 1991 until her death in August 2006. Dr. Shapiro provided valuable continuity to the Panel's deliberations while simultaneously offering a fresh perspective that was rooted in observations from her clinical practice and was supported and substantiated by her clinical research and indepth understanding of the literature. Dr. Shapiro had a passion for improving asthma care and an unwavering commitment to develop evidence-based recommendations that would also be practical. Dr. Shapiro inspired in others the essence of what NAEPP hopes to offer with this updated Expert Panel Report: a clear vision for clinicians and patients to work together to achieve asthma control.
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

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

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

We assert that without a measure of whether or not the reason the child is in the emergency room is sufficient to make it a clinically appropriate visit, it is impossible to interpret whether an ED visit represents a failure of clinical management and control, or a failure of the primary care or other aspects of the health care system to provide care at a more appropriate level of care.

Overarching statement: Even when not specifically indicated, we are interested in how these constructs are impacted by such factors as race, ethnicity, socioeconomic status or its indicators, or the presence of other special health care needs.

Our metric is designed to capture axes related to two distinct conceptual frameworks:

1) Asthma is a model of chronic disease management. In other words, ED visits may arise from acute exacerbations indicating a flare up of disease, and/or suboptimal management of the chronic illness.
2) ED visits for asthma may reflect limitations of primary care beyond the provision of suboptimal treatment, such as insufficient education, limitations of access or availability, breakdowns of communication, or a variety of other factors.

We note that the internal quality of the ED visit to manage the asthma is not the target of this measure. However, communication between the emergency department and the primary care site may prove to be within the scope of this measure, pending the views of our experts and developers.

Construct I: Need to sufficiently specify population for measure

Concept	Implications (Lay Statement)	Lit Review Questions
(Descriptive) **The measure will need to adequately specify the population that we consider to be eligible for an ED with asthma measure.**	The development of measures regarding ED use for children with asthma requires us to understand the strengths and weaknesses for our measure of various approaches to identifying whether or not children have asthma. It further requires us to understand the impact of the availability of various sources of data (such as encounter data, pharmaceutical data, electronic medical record or chart review data) on these strengths and weaknesses. We are aware that the use of the term asthma is variable. We are not interested in diagnoses with the name asthma, but with an operational diagnosis that we will functionally treat as asthma, whether it has been called chronic wheezing, reactive airway disease, chronic infectious bronchitis, etc. We recognize that asthma and its presentation may change over the course of a child's life.	1. When asthma care is evaluated, how is the population of care recipients defined? How is asthma defined? What is the impact of including various types of data (dx 1 or more, drugs, etc) on the sensitivity and specificity of asthma identification? What are practical and valid approaches to identifying asthma? How do the answers to these questions differ between adults and children? 2. Are any groups persistently excluded from studies of asthma care (i.e., are children who have asthma and other comorbid conditions, such as a malignant disease, excluded?). What rationale is provided for the exclusion? 3. Are any non-asthma diagnoses considered to be indicators of asthma or potential asthma (e.g. bronchitis, bronchiolitis, wheezing, atopy) 4. For children up to age 21, how do issues of diagnosis, management, and follow-up differ by age and developmental stage? 5. At what point does literature suggest that reactive airway disease should be managed as asthma? a. What other conditions are managed as asthma? 6. What common current or preexisting comorbid conditions alter the management plan for asthma?

Construct II: Adequacy of management of asthma (as a chronic disease example)

Concept	Implications (Lay Statement)	Lit Review Questions
IIA. **↑Adequacy of asthma management:** **↓ED visits**	Since asthma is a chronic disease characterized by acute exacerbations, the extent to which asthma care is optimized through the use of appropriate medications, the control of the	1. What are the recommendations of the NHLBI guidelines? a. What does the literature suggest about the usefulness of NHLBI guidelines? b. Are there aspects that it has identified that appear to be missed?

	environment, and the preparation of the parent/child dyad to adapt to changes in circumstances (e.g. viral respiratory infection or exposure to cold) should reduce the number of ED visits, irrespective of the number of primary care visits.	2. What do we know about asthma management, how it's measured, who provides it, patterns of care and how ED visits vary as a consequence? 3. Does identification of PCP improve outcomes of ED visit, including patterns of care, utilization? 4. What do we know about the content of an asthma plan and its relationship to a full program of chronic disease management, and its influence on ED utilization? 5. What evidence is there about the impact on outcomes such as ED use when the child or adolescent is involved in asthma self-management? For example, does it matter if: a. The child has a written asthma plan? b. The child understands their asthma plan? c. The child is given an opportunity to participate in managing care? 6. How is the role of the child in self-management measured? 7. How much are children able to recognize, communicate and act on their asthma? 8. What do we know about the impact of asthma services on asthma management? This includes: a. Treatment from an asthma specialist; b. Social worker; or c. Multidisciplinary personnel 9. To what extent is ED use by children with asthma stimulated by non-asthma related issues? a. How can we identify when that occurs? b. What is the evidence that providing other services will reduce the number of ED visits? 10. To what extent do children contribute to their management (including avoiding triggers, recognizing symptoms, medication adherence, etc.)? a. What is the impact and variance by age? 11. What is the evidence regarding adequacy of various medication delivery systems for infants, toddlers, children and adolescents in acute and chronic settings?

	12. Is there evidence of prior insult to the lungs such as sequelae of prematurity, etc. that create distinct subpopulations when considering this measure (at risk for ER visit)? 13. What aspects of the health services environment have been identified as contributing to outcomes of asthma management (e.g. school based health care)? 14. Does rate of ED utilization for non-respiratory diagnoses vary between asthmatics and non-asthmatics? 15. What is known about how often children with asthma use the ED over an extended period of time? Does it change over the life course of childhood? How does that vary by child characteristics, including race, SES, urban, suburban vs. rural, and age?

IIB. **↑PCP capacity/knowledge/skill:** **a. ↑Asthma management** **b. ↓Asthma exacerbations** **c. ↑Chronic disease management**	Broadly speaking, patient management of asthma is influenced by the capacity of the PCP practice. This includes the knowledge and skills possessed by the PCP, as well as office support to enhance access and availability of care. PCP includes the ability of the PC office to meet the cultural needs of the patient and their family.	1. What are the diversity of practices or services that may or may not impact ability or capacity of the PCP practice to manage asthma? 2. What do we know about the specific skills and processes that contribute to a primary care practice's capacity? 3. What patterns of visits or medication use or other indicators have been used as markers of well or poorly delivered primary care for asthma in children and/or adults? 4. What is the minimum use of specialists appropriate for children with asthma? How does that vary with history of ED or hospital use? a. When and how does the use of specialists become a marker for higher or lower quality of care? 5. What evidence is there regarding the nature of the PCP practice for children with asthma? For example, the level of continuity with individual clinicians vs. practices, the accessibility of specified clinicians and/or practices during the day and/or after work hours, etc.

IIC. **↑Asthma education:** **a. increases recognition of symptoms >** **b. ↑Management skills**	Enhancing what patients or their families know about asthma may be an important tool to improve care for children with asthma. The likely first effect of such education is to enhance the capacity of a caregiver to identify what symptoms may relate to asthma. This could conceivably increase utilization of both PCP and ED services if this were to increase the caregiver's perceived need for care for their child's asthma. With a more sophisticated understanding, including having a valid asthma action plan and understanding how to use it, ED care may be reduced and PCP care for asthma may be reduced, as symptoms are less frequent and parents are more competent to manage them when they arise.	1. What are metrics or processes regarding the quality of asthma care? Is it drug ratios (i.e. proportion of prescriptions filled that are for rescue vs control medications), asthma action plan, , capacity of PCP office, relationship to PCP practice, or other specific bundles of care, etc? 2. What constitutes "perfect care"/"best practice" for any specified type of patient? 3. What do we know about the impact of asthma education programs on quality of care, outcomes of care, or utilization of care? Define utilization of care as including: a. PCP utilization, b. ED utilization, c. Referral/specialist utilization, d. Non physician care team member utilization, e. Medication usage, f. Hospitalizations, and/or g. Other care utilization areas to consider? Examples may include functional status, quality of life elements, spirometry, role functioning. 4. What is the diversity of asthma education programs and what are the differences in quality of care/outcomes/utilization of care associated with differences? 5. Does referral to an asthma specialist impact quality of care, utilization of care and asthma outcomes? 6. Does referral to a social worker impact utilization of care and asthma outcomes? 7. (Broad) Does involvement of multidisciplinary personnel (beyond allopathic or osteopathic physicians) impact quality of care, utilization of care and asthma outcomes? 8. What are desirable roles and effectiveness of interventions that extend beyond the healthcare system, such as reducing pollution, focusing on environmental justice, housing, dust mites, etc.? 9. How does organization and capacity of the practice setting influence the delivery of asthma management education?

Construct III: **Adequacy of PCP practice site to handle acute exacerbations of chronic disease and/or acute illnesses**

Concept	Implications (Lay Statement)	Lit Review Questions
IIIA. **↑Primary care capacity:** **a. ↑ PCP visits (routine, WCC)** **b. ↑PCP visits (other acute dx)** **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)** **IIIA.2** **SUBCONSTRUCT:** **↑Accessibility:** **a. ↑ PCP visits (routine, WCC)** **b. ↑PCP visits (other acute dx)** **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)**	In general, enhanced capacity may affect a patient's access to care. Capacity can refer to patient services that make it easier for a patient to receive timely care, such as location or hours of offices, to the ability to triage phone calls in a timely and effective way, or may include the materials and services present within an office (e.g. the presence of a treatment room, the capacity to deliver oxygen, nebulizers, etc.) Such capacity may be limited or enhanced by staffing, space, the ability to safely transport someone from the office to a hospital, etc. If PCP office capacity is optimized, ED visits may be reduced as acute and mundane conditions can be managed in a PCP setting. Subsequently, increased capacity of the entire PCP support network will increase number of PCP visits.	1. What do we know about access to the PCP's office as a place to manage asthma, and the subsequent capacity of a PCP and the diversity of practice settings? Additionally, how do we measure capacity and, its impact on QoC, processes of care, asthma outcomes, asthma specific processes and utilization? How do these factors impact ED use or other outcomes? a. In general: i. PCP/specialist ratio in a plan or PCP/child ratio ii. PCP time spent in visit (incl. minutes per sick, well-child, asthma management visit) iii. Nature of training activities iv. How long does it take to schedule a visit (incl. asthma (chronic), acute, follow-up visit) v. Office hours and visit flexibility (incl. after hours coverage, office consult, meet in ED) vi. Phone capabilities: (incl. answering capacity, putting on hold, returning calls, after hours phone service) vii. Level of implementation of patient centered medical home/chronic care model, eg i. Use of registries ii. Standardized tools for measurement iii. Case management iv. Group visits or other education, etc b. Specifically, ability to manage acute dx in office, which includes: i. Do they have a treatment room or capacity to use a room as a treatment room? ii. Do they offer rescue treatments (e.g. nebulizers, spacers)? iii. Can they measure oxygen saturation? iv. Do doctors feel comfortable with acute asthmatic in office? v. Can they take time to manage an acute pt in their office?

vi. Do they have safe and rapid transport to a hospital (how long?)

2. Availability and accessibility of offices (incl. office hours, geographic distribution)
   a. What do we know about linguistic capabilities in the PCP setting influencing use of the ED?
   b. What do we know about proximity of the PCP office to public transit on the utilization of the ED?
3. What do we know about the impact of variations in patterns of care/practice, use of modalities, and/or and receipt of well-child care on asthma management or outcomes (eg ED use)? Does Immunization status reflect on t eh capacity of the PCP, on the state of the child, or on other factors that may relate to asthma outcomes? How about the sufficiency of the number of WCC Visits (eg meets HEDIS standard or AAP standard or does not)? Absolute number of visits to PCP?
4. Are children with more WCC visits less likely to use the ED for acute visits? children who are UTD on their immunizations?
5. What literature is there on the relationship between pediatric ED use and other measures of asthma exacerbation/outcomes?
6. What do we know about variability of capacity and management of mundane conditions (e.g. OM, URIs, pharyngitis), office to ED ratios?
7. What do we know about variability of capacity and management of acute conditions requiring interventions (e.g. asthma)?
8. To what extent does ED capacity increase use of ED services? Do hospitals advertise ED services, have fast track for mundane conditions, etc?
9. To what extent does ED have capacity to provide primary care, routine immunizations, etc? How is that built into policies and protocols?
10. At what age does the PCP start meeting alone with child? Time spent in visit?
11. To what extent and at what age do PCP's involve children in self-management and does it vary?

IIIB. **↑Relationship with PCP:** **a. ↑ PCP visits (routine, WCC)** **b. ↑PCP visits (other acute dx)** **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)**	Improved relationship with PCP may increase visits to your PCP and decrease ED visits, for both acute and mundane conditions. A good relationship may lead to greater trust and adherence to recommendations (both WCC and asthma care) and drive a preference for seeking care by the PCP over seeking care in another environment. In general, we are referring to relationship of caregiver with PCP and their office staff. We recognize the importance of the relationship of PCP's with patients as well; when the relationship between the PCP and the child rather than caretaker is emphasized in research, we'd like to capture that as well.	1. What exists regarding measuring the quantity and quality of the relationship with PCP? Specifically: a. What's the variation and does it matter? b. How is it measured? c. What do we know about patient experience of care, especially as it relates to relationship with clinicians/PCP d. To what extent is quality of relationship expressed in terms of caregiver vs. child relationships and how does this change with age of child or longevity of connection to a PCP? 2. What evidence is there regarding use of supplemental services outside of regular clinical visits and how do these services impact quality and utilization of care? Define supplemental services as: a. Electronic educational/reminder tools (incl. social media) b. Telephone educational/reminder tools c. Print materials (e.g. educational brochures) d. Disease management, demand management, or other type programs e. Other services to consider? Measure quality, utilization of care should include at least : a. ED visits b. PCP visits 3. How does role of child in self care/management tie into these issues?

Construct IV: The connectedness of care in the primary care and ED setting – before, during, and after of the ED visit

Concept	Implications (Lay Statement)	Lit Review Questions
IV. (Descriptive) Enhanced integration of ED care of asthma with routine care will have better outcomes	If primary care is generally pretty good, then the ED visit should be an extraordinary event. In such cases the PCP alerting the ED to current management and the ED assuring appropriate follow up with the PCP is important. In cases where primary care is of lower quality or more variable, the ED visit may enhance the long term management of the child with asthma. And we need to assess this. One of the ways it might do so is to construct an asthma management plan that is then followed by the PCP. Another way is to connect a child without adequate primary care to primary care, especially to someone who is competent to manage the asthma.	1. What evidence supports that ED visits for asthma are most effective when visit is followed by a visit to the PCP? 2. Do utilization patterns in both the ED and primary care setting change following ED visits? 3. Is an effective/more effective use of medications seen following an ED visit? 4. Does the identification of a primary care provider improve outcomes of an ED visit (including patterns of care utilization)? 5. Is pre or intra visit communication with the primary care provider associated with better outcomes? How often does this occur? Are there systematic differences regarding those for whom this does and does not occur? 6. Are ED visits for asthma routinely associated with some form of communication or linkage with PCP? Does that result in better outcomes?

Construct V: Equity is a value in asthma care

Concept	Implications (Lay Statement)	Lit Review Questions
V. (Descriptive) Equity is a critical construct of quality for children with equity	Systematic differences in the frequency or nature of ED visits for asthma on the basis of race, ethnicity, family make-up, income/economic status, specifics of insurance status, presence or absence of comorbid special health care needs, etc represents decrements in quality that our measures should identify.	1. Does the literature indicate systematic or predictable differences in the frequency or nature of asthma care for children as it relates to ED visits for asthma that may be interpreted as representing inequitable structures, processes, outcomes, experiences with, or coordination of care? 2. What do we know about how social determinants and diagnosis and management of asthma and its outcomes, specifically as it relates to use of ED? 3. What do we know about the extent to which use of the ED for children with asthma that relates to the external physical and social environment?

Proposed Research Questions

Asthma- We propose to prioritize our Asthma Construct Table, to the following questions:

Acronyms

PCP: Primary Care Provider

ED: Emergency Department

WCC: Well-child care

Baseline Question (for Questions 1, 2 and 3 below):

When asthma care is evaluated, how is the population of a t the population level? What are specific implications of how you ident ng various approaches to specifying the denominator of children l and valid approaches to identifying asthma at the population level? How do the answers to these questions differ between adults and children?

Question 1 (Construct IIA.2):

For children with asthma, what do we know about asthma management? How is management of asthma described and measured? This includes who (PCP, asthma specialist, ED, etc) primarily manages it as well as who provides it. What are the patterns of care and what do we know about how use of the ED varies as a result of various approaches to management?

- **Question 1a (Construct IIB.3):**

Specifically, have any of these patterns of visits or medication use or other characteristics of care been used as markers of well or poorly delivered primary care for asthma for children and/or adults?

Question 2 (Construct IIB.5):

How has varying asthma care for children been described on the basis of characteristics of the PCP offices or practices? For example, are they characterized by the level of continuity between individual clinicians, the level of conntiuity with any provider in the practice, the accessibility of specified clinicians and/or practices during the day and/or after work hours, etc?

- **Question 2a (Construct IIIA.3):**

What do we know about the impact of variations in patterns of care/practice, use of treatment modalities, and/or receipt of well-child care on asthma management or outcomes (e.g. ED use)? How about the sufficiency of the number of WCC Visits (eg meets HEDIS standard or AAP standard or does not)? Absolute number of visits to PCP?

Question 3 (Construct IIC.7):

(Broad) Does involvement of multidisciplinary personnel (beyond allopathic or osteopathic physicians) impact quality of care, utilization of care and asthma outcomes both within context of a primary care practice or in other clinical settings?

- **Question 3a. (Construct IIIB.2):**

What evidence is there regarding use of supplemental services outside of regular clinical visits and how do these services impact quality and utilization of care?

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

We conducted a scoping review as follows:
We identified key constructs of asthma ED use measures for consideration. We created a table of these constructs in technical and lay language, and listed research questions for the review to answer. Our contractor (a national accrediting body experienced in measure development), prepared for us a literature review in 2 stages and we supplemented this with targeted reviews as needed to answer specific questions that arose during the measure development process.

The above construct table was used to guide the review and was the basis for the first round of review. Following the table, we include a list of questions for focused review that guided round 2 of the review, which resulted in a detailed summary of 91 articles from the peer-reviewed literature. In addition to this review, the CAPQuaM scientific team conducted an ad hoc series of reviews to answer specific questions such as the reliability of administrative data to identify asthma, and the value of expert panels and the RAND/UCLA appropriateness method. The CAPQuaM degree 360 method starts with a topic area and the measures emerge during the process, in this case necessitating the specified ad hoc reviews.

We searched peer reviewed and gray literature from 1985-2014 over the course of these reviews. Literature was summarized for our expert panel, which met in late 2013.

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

Our approach to developing this measure stems from several vibrant and scientifically sound traditions. We first discuss research involving the soundness of our data sources, which include both administrative data to identify cases (and a fraction of numerator qualifications) and chart review (medical record audit) to confirm some denominator inclusions and to identify most numerator inclusion. This is a generally accepted and standard approach with acceptable reliability.

Brook and Davies [1] trace the early history of quality measurement and remind us of the importance of medical chart audit as an approach to quality measurement. Lohr and Brook at RAND and Roos in Manitoba, Canada pioneered the use of electronically-available administrative data (generated by routine health care operations, such as billings) as proxies for health care processes. Administrative data carefully used reduces burden of quality measurement. [2-6]

As the National Committee for Quality Assurance (NCQA) developed the Healthcare Employee Data Information Set (HEDIS) as the de facto measurement system for managed care, attention turned to the use of administrative data for routine performance measurement. Research demonstrated that administrative data could have a role in producing quality measures, with augmentation by chart review often necessary. Administrative data are not typically sufficient for detailed clinical assessment. [7-11] HEDIS developed a hybrid approach, using administrative data and chart review, which this measure borrows heavily from. [12, 13]

We have used rigorous and transparent methods [14] to assemble a national expert panel that included pediatricians, family physicians, pediatric and general emergency room specialists, a pediatric pulmonologist and a pediatric allergist from practices and medical schools around the country. This work was conducted in collaboration with national clinical societies (AAP, AAFP) and CAPQuaM's diverse other partner organizations, including NY State DoH/Medicaid. NCQA is an important technical consultant and partner. The specific criteria that we operationalize in this

measure were all rated by the expert panel with a median score of 8 or 9 on a 9 point scale (9 high) as circumstances for which the ED is an appropriate level of care. The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children.

Select references documenting other aspects of performance gap, and supporting our process and data sources are also noted (15-35).

1. Brook, R.H. and A. Davies-Avery, *Mechanisms for Assuring Quality of US Medical Care Services: Past, Present, and Future*. 1977: Santa Monica, CA.
2. Brook, R.H. and K.N. Lohr, *Efficacy, effectiveness, variations, and quality. Boundary-crossing research.* Med Care, 1985. **23**(5): p. 710-22.
3. Lohr, K.N., *Use of insurance claims data in measuring quality of care.* Int J Technol Assess Health Care, 1990. **6**(2): p. 263-71.
4. Roos, L.L., et al., *Using Administrative Data Banks for Research and Evaluation: A Case Study.* Evaluative Research 1979. **236**(3).
5. Roos, L.L., Jr., J.P. Nicol, and S.M. Cageorge, *Using administrative data for longitudinal research: comparisons with primary data collection.* J Chronic Dis,1987. **40**(1): p. 41-9.
6. Roos, L.L., Jr., et al., *How good are the data? Reliability of one health care data bank.* Med Care, 1982. **20**(3): p. 266-76.
7. Dresser, M.V., et al., *Clinical quality measurement. Comparing chart review and automated methodologies.* Med Care, 1997. **35**(6): p. 539-52.
8. Newton, K.M., et al., *The use of automated data to identify complications and comorbidities of diabetes: a validation study.* J Clin Epidemiol, 1999. **52**(3): p. 199-207.
9. Thompson, B.L., et al., *Measuring clinical performance: comparison and validity of telephone survey and administrative data.* Health Serv Res, 2001. **36**(4): p. 813-25.
10. Angier, H., et al., *Variation in outcomes of quality measurement by data source.* Pediatrics, 2014. **133**(6): p. e1676-82.
11. Weiskopf, N.G. and C. Weng, *Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research.* Journal of the American Medical Informatics Association, 2013. **20**(1): p. 144-151.
12. Pawlson, L.G., S.H. Scholle, and A. Powers, *Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures.* Am J Manag Care, 2007. **13**(10): p. 553-8.
13. NCQA. *National Committee for Quality Assurance*. [cited 2014 7/30/14]; Available from: http://www.ncqa.org/.
14. Brook, R.H., et al., *A method for the detailed assessment of the appropriateness of medical technologies.* International journal of technology assessment in health care, 1986. **2**(01): p. 53-63.
15. DeAngelis, C., P. Fosarelli, and A.K. Duggan, *Use of the emergency department by children enrolled in a primary care clinic.* Pediatr Emerg Care, 1985. **1**(2): p.61-5.
16. Talreja, N., et al., *Modifiable factors associated with severe asthma exacerbations in urban patients.* Ann Allergy Asthma Immunol, 2012. **109**(2): p.128-32.
17. Auger, K.A., et al., *Medical home quality and readmission risk for children hospitalized with asthma exacerbations.* Pediatrics, 2013. **131**(1): p. 64-70.
18. Ducharme, F.M., et al., *Written action plan in pediatric emergency room improves asthma prescribing, adherence, and control.* Am J Respir Crit Care Med, 2011. **183**(2): p. 195-203.

19. Farber, H.J., *Optimizing maintenance therapy in pediatric asthma.* Curr Opin Pulm Med, 2010. **16**(1): p. 25-30.
20. Smith, S.R., D.B. Wakefield, and M.M. Cloutier, *Relationship between pediatric primary provider visits and acute asthma ED visits.* Pediatr Pulmonol, 2007. **42**(11): p. 1041-7.
21. Parchman, M.L. and S. Culler, *Primary care physicians and avoidable hospitalizations.* J Fam Pract, 1994. **39**(2): p. 123-8.
22. Prevention, C.f.D.C.a. *Home-based Multi-trigger, Multi-component interventions*. 2013 [cited 2013 May 20]; Available from: http://www.cdc.gov/asthma/interventions.htm.
23. Price, J.H., et al., *Racial/ethnic disparities in chronic diseases of youths and access to health care in the United States.* Biomed Res Int, 2013. **2013**: p. 1-12; Available in open access http://dx.doi.org/10.1155/2013/787616
24. Homer, C.J., et al., *Does quality of care affect rates of hospitalization for childhood asthma?* Pediatrics, 1996. **98**(1): p. 18-23.
25. Finkelstein, J.A., et al., *Quality of care for preschool children with asthma: the role of social factors and practice setting.* Pediatrics, 1995. **95**(3): p. 389-94.
26. Flores, G., et al., *Keeping children out of hospitals: parents' and physicians' perspectives on how pediatric hospitalizations for ambulatory care-sensitive conditions can be avoided.* Pediatrics, 2003. **112**(5): p. 1021-30.
27. Knudson, A., et al., *Disparities in pediatric asthma hospitalizations.* J Public Health Manag Pract, 2009. **15**(3): p. 232-7.
28. Bindman, A.B., et al., *Preventable hospitalizations and access to health care.* Jama, 1995. **274**(4): p. 305-11.
29. National Heart, Lung and Blood Institute. *Asthma Guidelines*. 2011 February 2011 [cited 2014 7/30/2014].
30. *A Matter of Urgency: Reducing Emergency Department Overuse* in *A NEHI Research Brief* 2010, New England Healthcare Institute.
31. Martin, B.C., *Emergency medicine versus primary care: a case study of three prevalent, costly, and non-emergent diagnoses at a community teaching hospital.* J Health Care Finance, 2000. **27**(2): p. 51-65.
32. Finkelstein, J.A., et al., *Comparing asthma care for Medicaid and non-Medicaid children in a health maintenance organization.* Arch Pediatr Adolesc Med, 2000. **154**(6): p. 563-8.
33. Owens, P.L., et al., *Care of children and adolescents in U.S. hospitals*, in *HCUP Fact Book No. 4*. Agency for Healthcare Research and Quality: Rockville: MD.
34. Pearson, W.S., et al., *State-based Medicaid costs for pediatric asthma emergency department visits.* Prev Chronic Dis, 2014. **11**: p. E108.
35. Taubman, S.L., et al., *Medicaid Increases Emergency-Department Use: Evidence from Oregon's Health Insurance Experiment.* Science, 2014. **343**(6168): p. 263-268.

The appropriate use criteria were derived from a set developed by an expert panel who synthesized the literature and their expert opinion into explicit criteria using the RAND/UCLA appropriateness method. Criteria that were rated 8 or 9 by the panel were included for this measure. The criteria set includes:

1) Hospitalization directly from the ED;
2) Documented physical findings consistent with respiratory distress, including:
    a) Labored breathing with retractions and/or evidence of accessory muscle use;
    b) Markedly decreased breath sounds;
3) O_2 saturation level less than 90 percent on percutaneous assessment;
4) An ABG obtained (or ordered);

5) Consultation ordered and obtained with a pulmonologist asthma specialist, an order of an arterial blood gas (ABG), or a consult with a pulmonary or asthma specialist.
6) Parent/caregiver referred to the ED after evaluation from the PCP or other office/clinic;
7) Parent/caregiver report of administering two or more doses of inhaled rescue medications without meaningful clinical improvement;
8) Parent/caregiver report that the child was in a pre-defined "red zone" of peak flow measurement as part of an asthma action or similar plan; or,
9) Parent/caregiver report of a rapid and life-threatening deterioration after a similar prior episode. This criterion is not included in the specifications for this measure.

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

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

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form
nqf_evidence_attachment_12_11_16.docx
1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?
Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

1b. Performance Gap

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

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

1b.1. **Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*
IF a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)
IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.
Asthma is one of the most common indications for emergency department (ED) visits by children. (1-3) AHRQ's Healthcare Cost and Utilization Project (HCUP) data from the Nationwide Emergency Department Sample (NEDS) found that in 2012, children between 1 and 17 years old had more than 1,895,000 ED visits for asthma with almost 10% resulting in hospitalization.

Evidence suggests that ED visits and hospitalizations in children with asthma vary systematically by how well-equipped that community is to provide primary care, and by the quality of primary care delivered. (4, 5) There is widespread literature illustrating that ED visits and hospitalizations are each undesirable utilization outcomes from poorly managed asthma. There is not a large literature that assesses whether or not pediatric ED visits were appropriate. (6 -10)

A body of literature has explored the value and feasibility of measuring the appropriateness of medical activities using data available in the medical record. (11-14) Early work in adults included assessment of hysterectomy, carotid endarterectomy and cardiac interventions. An independent research project brought the construct of appropriateness to children (15), while Kleinman and colleagues were the first to assess the appropriateness of specific pediatric procedures. (16, 17) A later study demonstrated the feasibility of medical record data for such an assessment. (18) DeAngelis pioneered studies of what constitutes a good reason to use the ED. (6) All of these studies used a definition of appropriateness that compared benefit to likely risk without specific consideration of costs. The need for more studies looking for overuse was recently reviewed. (19) RAND type Delphi panels are accepted around the world as a method for developing criteria to assess appropriateness. (20-22)

Research demonstrates that:

•ED visits are an important issue for child health insurers, including Medicaid, with clinical and financial consequences;

•An overcrowded primary care system contributes to ED use for non-emergent and even non-urgent conditions.

•Pediatric hospitalizations for asthma vary by primary care availability and quality

•ED visits are common for children with asthma, including those in Medicaid

•Assessment of appropriateness using information in the medical record is a well-established and validated method that has been successfully applied to children.

The literature suggests that a measure that assesses whether or not the ED is an appropriate level of care for a child with asthma at the time that they present has intrinsic value. Such a measure would:

•Characterize the process of care in a way that assesses whether a particular ED visit represents overuse

•Allow the outcomes of asthma care to be better characterized in a manner that describes performance and promotes targeted improvement. Inappropriate ED visits represent failures of primary care delivery, availability and/or access. Appropriate visits may represent a failure to control asthma. These have distinct and distinguishable meanings that contribute to the understanding of the quality of asthma care.

•Measuring the quality of asthma care requires assessment of multiple factors. This appropriateness measure helps plans, purchasers, and society to understand the implication of asthma ED visits as outcomes of asthma care. The implications herein is that understanding what is better or worse care requires looking at various factors and not simply a higher or lower appropriateness score. The understanding of this measure is enhanced by considering whether the rate of undesirable outcomes (ED visits and hospitalizations) is high or low and whether other measures of primary care availability and access or asthma quality suggest high levels of performance or not..

An abstract describing the proposed measure was peer-reviewed and subsequently presented to a national audience at AcademyHealth 2014 Annual Research Meeting in San Diego in the “Measuring the Safety, Quality, and Value” section. Feedback was positive regarding the methods, measures, ethics, and importance of this measure.

Research evidence supports the importance and need for our proposed measure that assesses whether the ED represents an appropriate level of care for children with asthma who are seen in the ED.

1.Kharbanda, A.B., et al., Variation in resource utilization across a national sample of pediatric emergency departments. J Pediatr, 2013. 163(1): p. 230-6.
2.Adams, J.G., Emergency department overuse: Perceptions and solutions. JAMA, 2013. 309(11): p. 1173-1174.
3.Institute, N.E.H., A Matter of Urgency: Reducing Emergency Department Overuse. Research Brief, 2010(March).
4.Perrin, J.M., et al., Variations in rates of hospitalization of children in three urban communities. N Engl J Med, 1989. 320(18): p. 1183-7.
5.Perrin, J.M., et al., Primary care involvement among hospitalized children. Arch Pediatr Adolesc Med, 1996. 150(5): p. 479-86.
6.DeAngelis, C., P. Fosarelli, and A.K. Duggan, Use of the emergency department by children enrolled in a primary care clinic. Pediatr Emerg Care, 1985. 1(2): p.61-5.

7.Berns, S.D., et al., Appropriate use of a pediatric emergency department: is the pediatrician called before the visit? Pediatr Emerg Care, 1994. 10(1): p. 13-7.
8.Rudowitz, R., A Look At CBO Projections For Medicaid and CHIP, in The Kaiser Commission on Medicaid and the Uninsured. 2014, The Henry J. Kaiser Family Foundation
9.Taubman, S.L., et al., Medicaid Increases Emergency-Department Use: Evidence from Oregon´s Health Insurance Experiment. Science, 2014. 343(6168): p. 263-268.
10.Smulowitz, P.B., et al., Increased Use of the Emergency Department After Health Care Reform in Massachusetts. Ann Emerg Med, 2014.
11.Brook, R.H., et al., A method for the detailed assessment of the appropriateness of medical technologies. Int J Technol Assess Health Care, 1986. 2(1): p. 53-63.
12.Park, R.E., et al., Physician ratings of appropriate indications for six medical and surgical procedures. Am J Public Health, 1986. 76(7): p. 766-72.
13.Fitch, K., et al., The RAND/UCLA Appropriateness Method User´s Manual. 2001 RAND.
14.Kosecoff, J., et al., The appropriateness of using a medical procedure. Is information in the medical record valid? Med Care, 1987. 25(3): p. 196-201.
15.Kemper, K.J., Medically inappropriate hospital use in a pediatric population. N Engl J Med, 1988. 318(16): p. 1033-7.
16.Kleinman, L.C., et al., The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States. Jama, 1994. 271(16): p. 1250-5.
17.Kleinman, L.C., E.A. Boyd, and J.C. Heritage, Adherence to prescribed explicit criteria during utilization review. An analysis of communications between attending and reviewing physicians. Jama, 1997. 278(6): p. 497-501.
18.Keyhani, S., et al., Electronic health record components and the quality of care. Med Care, 2008. 46(12): p. 1267-72.
19.Keyhani, S. and A.L. Siu, The underuse of overuse research. Health Serv Res, 2008. 43(6): p. 1923-30.
20.Bernstein, S.J., et al., The appropriateness of hysterectomy. A comparison of care in seven health plans. Health Maintenance Organization Quality of Care Consortium. Jama, 1993. 269(18): p. 2398-402.
21.Taylor, A.J., et al., ACCF/SCCT/ACR/AHA/ASE/ASNC/NASCI/SCAI/SCMR 2010 Appropriate Use Criteria for Cardiac Computed Tomography. A Report of the American College of Cardiology Foundation Appropriate Use Criteria Task Force, the Society of Cardiovascular Computed Tomography, the American College of Radiology, the American Heart Association, the American Society of Echocardiography, the American Society of Nuclear Cardiology, the North American Society for Cardiovascular Imaging, the Society for Cardiovascular Angiography and Interventions, and the Society for Cardiovascular Magnetic Resonance. J Cardiovasc Comput Tomogr, 2010. 4(6): p. 407.e1-33.
22.Basger, B.J., T.F. Chen, and R.J. Moles, Validation of prescribing appropriateness criteria for older Australians using the RAND/UCLA appropriateness method. BMJ Open, 2012. 2(5).

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

In testing we found that for children age 2-5, 181 of 335 (54.3%) were deemed appropriate, with the breakdown of reasons for appropriateness presented in the Testing Form 2b.2.2. Other age groups found that children 6-11, 209 of 447 (43.8%) ED visits were appropriate, while for adolescents aged 12-18, 165 of 341 (48.4%) visits were appropriate. These numbers were sufficient to identify statistically significant differences in the proportion that were appropriate between age groups, among racial/ethnic groups, and within age group among racial/ethnic groups. These data demonstrate that the specified sample size is sufficient to find meaningful differences between groups at the various specified levels. In our work, validating and testing the measure for the rate of appropriateness, we have demonstrated the capacity to identify the included events (ED visits and hospitalizations) using administrative data and our specifications for identifiable asthma. That aspect of testing was conducted using state wide data from the NY State Medicaid Managed Care Program.

We also incorporate by reference work done by our partner NCQA that demonstrate the capacity of administrative data to identify a population with asthma.

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

In her seminal article nearly three decades ago, DeAngelis included an asthma attack as an appropriate indication for use of the ED.[1] As a common chronic illness characterized by remissions and potentially preventable exacerbations undesirable utilization outcomes for asthma have been a frequent target for measurement for three decades. Reducing the relative number of ED visits during the care for asthmatic children remains a high priority on the national agenda. The universal delivery of optimal asthma care has the potential to lower costs and improve quality of life. Understanding which ED visits represent failures of clinical prevention and which instead represent a mismatch of service level to clinical need can help to move these goals forward. The submitted measure is a step in this direction.

ED visits for asthma can be reduced through both enhanced access to care and through better quality of care. The NIH's National Asthma Education and Prevention Program Guideline has been shown to reduce the frequency of breakthrough asthma and of ED visits and hospitalizations when implemented. The literature points to two general characteristics of asthma care delivery systems that correlate with ED utilization. One is the effective use of preventive and routine care measures, such as multidisciplinary practice or a medical home model, the presence of an asthma action plan, the use of controller medications supplemented by judicious use of rescue medications. [2-6] The other is the availability of primary care or urgent care visits as a step before ED use in the context of either a general pediatric or an asthma specialty practice. [6, 7] Conversely, a lack of comprehensive asthma care, which includes primary and secondary prevention schemas, and a lack of available urgent care services are both commonly cited as reasons for preventable ED visits. It has been demonstrated that the children who used the ED underutilized primary care services [6] and it has also been demonstrated that interventions that attempt to provide comprehensive, multidisciplinary care are able to decrease ED utilization for asthma care.[8] We acknowledge that environmental management and control is a nonclinical opportunity to improve the quality of life for children with asthma and to reduce health care utilization, but do not focus on these issues in this submission.

High rates of asthma visits to the ED suggest widespread deficiencies in asthma care. The literature shows that lack of proper asthma care is disparate with minority children bearing undue burden. [9-11]

The literature also presents different perspectives on appropriate use of the ED for pediatric asthma. Pediatric asthma is one of the leading conditions when it comes to potentially avoidable ED visits. [12] Asthma has been classified both as an avoidable hospitalization condition (AHC) and as an ambulatory care sensitive condition. This describes that a meaning proportion of ED visits or hospital admissions could have been avoided with proper outpatient care. [12, 13] Poor outpatient care can be an outcome of a number of variables. As noted, the availability of primary care can reduce such inappropriate and costly visits. [7, 12, 14 -17]

Assessing the extent to which ED use for asthma is appropriate can inform health policy, manpower planning, and clinical quality improvement activities. It can help to answer the question of how much of ED use potentially may be prevented by better management of the underlying asthma, versus how much requires other, process or structural improvements to reduce use of the ED when a lower level of care would meet the clinical needs of the child. Refractory asthma or those with unavoidable environmental exposures leading to an acute exacerbation requiring medical care are likely to be identified as appropriate, reminding us that NOT all asthma ED visits are preventable even with optimal care.

With a better understanding of ED use, health care organizations and policy makers could develop better informed approaches to optimizing services for children with asthma. And hopefully children and their families may increasingly be spared the inconvenience, risk, and costs of ED visits for asthma.

References:

1. DeAngelis, C., P. Fosarelli, and A.K. Duggan, Use of the emergency department by children enrolled in a primary care clinic. Pediatr Emerg Care, 1985. 1(2): p.61-5.
2. Talreja, N., et al., Modifiable factors associated with severe asthma exacerbations in urban patients. Ann Allergy Asthma Immunol, 2012. 109(2): p.128-32.
3. Auger, K.A., et al., Medical home quality and readmission risk for children hospitalized with asthma exacerbations. Pediatrics, 2013. 131(1): p. 64-70.
4. Ducharme, F.M., et al., Written action plan in pediatric emergency room improves asthma prescribing, adherence, and control. Am J Respir Crit Care Med, 2011. 183(2): p. 195-203.
5. Farber, H.J., Optimizing maintenance therapy in pediatric asthma. Curr Opin Pulm Med, 2010. 16(1): p. 25-30.
6. Smith, S.R., D.B. Wakefield, and M.M. Cloutier, Relationship between pediatric primary provider visits and acute asthma ED visits. Pediatr Pulmonol, 2007. 42(11): p. 1041-7.
7. Parchman, M.L. and S. Culler, Primary care physicians and avoidable hospitalizations. J Fam Pract, 1994. 39(2): p. 123-8.
8. Prevention, C.f.D.C.a. Home-based Multi-trigger, Multi-component interventions. 2013 [cited 2013 May 20]; Available from: http://www.cdc.gov/asthma/interventions.htm.
9. Price, J.H., et al., Racial/ethnic disparities in chronic diseases of youths and access to health care in the United States. Biomed Res Int, 2013. 2013: p. 1-12; Available in open access http://dx.doi.org/10.1155/2013/787616
10. Homer, C.J., et al., Does quality of care affect rates of hospitalization for childhood asthma? Pediatrics, 1996. 98(1): p. 18-23.
11. Finkelstein, J.A., et al., Quality of care for preschool children with asthma: the role of social factors and practice setting. Pediatrics, 1995. 95(3): p. 389-94.
12. Flores, G., et al., Keeping children out of hospitals: parents´ and physicians´ perspectives on how pediatric hospitalizations for ambulatory care-sensitive conditions can be avoided. Pediatrics, 2003. 112(5): p. 1021-30.
13. Knudson, A., et al., Disparities in pediatric asthma hospitalizations. J Public Health Manag Pract, 2009. 15(3): p. 232-7.
14. Bindman, A.B., et al., Preventable hospitalizations and access to health care. Jama, 1995. 274(4): p. 305-11.
15. National Heart, Lung and Blood Institute. Asthma Guidelines. 2011 February 2011 [cited 2014 7/30/2014].
16. A Matter of Urgency: Reducing Emergency Department Overuse in A NEHI Research Brief 2010, New England Healthcare Institute.
17. Martin, B.C., Emergency medicine versus primary care: a case study of three prevalent, costly, and non-emergent diagnoses at a community teaching hospital. J Health Care Finance, 2000. 27(2): p. 51-65.

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

Race/Ethnicity

Our medical chart audit found that the measure varies by race/ethnicity. Hispanic children had higher rates of questionable use of the ED (55.9% of visits) when compared to non-Hispanic children (46.8%), p=.002. Black children showed a trend toward more questionable use compared to all other children (53.6% questionable vs 48.7%, p=.10). Overall, Blacks had an appropriate use rate of 51.3%, Whites 56.5%, Hispanics 44.1% and other races 45.2%. For ages 2-5, Blacks had an appropriate use rate of 57.1%, Whites 63.6%, Hispanics 50.9%, and other races 51.9%. For ages 6-11, Blacks had an appropriate use rate of 49.3%, Whites 50.0%, Hispanics 46.3% and other races 39.8%. For ages 12-18, Blacks had an appropriate use rate of 49.3%, Whites 66.7%, Hispanics 46.3% and other races 46.7%. Chi-square analysis confirms that these differences are statistically significant.

Insurance Status
Overall, the appropriate use rate for Medicaid patients was 46.3%, Private insurance 59.0%, Uninsured patients 38.6% and other forms of insurance (military and Worker's comp) 55.0% (p=.005). Within the age strata, for ages 2-5 the appropriate use rate for Medicaid patients was 53.9%, Private 67.4%, Uninsured 40.9% and other was 20%. For ages 6-11, the appropriate use rate for Medicaid patients was 41.5%, Private 57.7%, Uninsured 35.7% and other was 52.6%. For ages 12-18, the appropriate use rate for Medicaid patients was 46.1%, Private 54.5%, Uninsured 42.1% and other was 68.8%. Chi-square analysis demonstrates the presence of statistically significant differences.

Socioeconomic Status
The measure is specified to be stratified in 2 ways to assess aspects related to socioeconomic status: Public versus Commercial Insurance, and by 5 strata defined by the percent of the population in poverty in their county of residence.

Rurality/Urbanicity
These measures are specified to be reported by Urban Influence Codes (UIC), which have been developed by the USDA based on a number of criteria to describe the levels of urbanicity and rurality. This is intended not only to report within plan differences but to allow for aggregation as appropriate. While each UIC has its own meaningful definition, some researchers choose to aggregate various codes. We recommend consideration of the aggregation schema of Bennett and colleagues at the South Carolina Rural Research Center. (2) Their aggregation scheme brings together Codes 1 & 2 as Urban; 3, 5, & 8 as micropolitan rural; 4, 6, & 7 as rural adjacent to a metro area; and 9, 10, 11, & 12 as remote rural. We observe that UIC 5 might as well be aggregated with 4, 6, & 7 as an adjacent rural area. Further, while this approach to rurality does not map exactly to the population density based definition of frontier (< 6 persons per square mile) as articulated in the Affordable Care Act, use of such categories is consistent with the ACA's intent that the Secretary ask that data that are collected for racial and ethnic disparities also look at underserved frontier counties. Frontier health care may be approximated by analysis of the remote rural categories. (3)

This judgment was confirmed after CAPQuaM consulted with Gary Hart, Director of the Center for Rural Health at the University of North Dakota School of Medicine & Health Sciences, who is heading a HRSA-funded project to develop new methods to analyze frontier health. We clarified that his work suggests that UIC 9-12 is the best overall approach to using county level data to study frontier health. Inclusion of UIC 8 would make the analysis more sensitive to including frontier areas but at a meaningful cost in sensitivity.

Those interested in care specific to large cities may wish to aggregate the rural area and analyze UIC 1 and 2 separately. Frontier health care may be approximated by analysis of the remote rural categories. (3) The New York State Medicaid data were sensitive to urbanicity with higher rates of ED utilization in the most urban areas and lowest in the most rural areas and other areas intermediate between the two.

For aggregation and as an imperfect approximation one can also group as urban (1 and 2), suburban (3-6) and rural (7-9). This is what we have used for our NY Medicaid analysis to demonstrate that variations are observed for this measure using UIC codes.

1.Kawachi I, B.L., Neighborhoods and Health. 2003, New York, NY: Oxford University Press.
2.Bennett, K.J., Olatosi B. & Probst, J.C., Health Disparities: A rural-urban chartbook. 2008, Columbia, South Carolina: South Carolina Rural Health Research Center.
3.Hart, G., Frontier/Remote, Island, and Rural Literature Review. 2012.

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

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

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

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
We will create a webpage as soon as possible, likely early in the new year.

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Asthma_5_12_06_16.xlsx

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

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

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator is the number of eligible asthma ED visits in the random sample that also satisfy at least one of the explicit criteria to indicate that the ED is an appropriate level of care. Distinct numerators are reported for children ages 2-5, 6-11, 12-18, and optionally, 19 - 21.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Children and adolescents who have a qualifying ED visit associated with asthma as the first or second diagnosis and are in the random sample;

AND have at least one of the following:

•Disposition of the ED visit was admission to the hospital,
OR
•Documented physical findings consistent with respiratory distress, including any of the following:
o Labored breathing (including moderate or severe increased work of breathing); OR
o Retractions, grunting, and/or evidence of accessory muscle use; OR
o Markedly decreased breath sounds;
OR
•Recorded oxygen saturation below 90%;
OR
•An arterial blood gas (ABG) was obtained in the emergency department;
OR
•The child had a consultation with a pulmonologist or asthma specialist that was ordered and provided in the ED;
OR
•There is clear documentation that prior to arrival in the ED any of the following occurred:
o The child was referred to the ED after evaluation by the PCP or other clinician. The evaluation may include an in person visit or auscultation including via telephone. OR
o The child received two or more doses of inhaled rescue medications without sufficient clinical improvement. Documentation of parent report meets the criterion. OR
o The child was assessed with an objective instrument such as a peak flow meter and was found to be in a pre-defined "red zone" of peak flow measurement as part of an asthma action or similar plan. Documentation requires ALL of the following: a written asthma action plan exists AND defines a "red zone" for which urgent assessment by a clinician is indicated AND an objective assessment was made and its result was in the pre-defined red zone. Documentation of parent report meets the criterion.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator is a random sample of the patients in each age stratum who have visited the emergency department for asthma (as a first or second diagnosis) and meet the specified criteria for having identifiable asthma (defined in s2b).

Separate numerators and denominators are reported for children age 2-5, 6-11, 12-18, and, optionally, 19-21 years. An overall rate across strata is not reported.

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Denominator Elements:

The presence of identifiable asthma (see table 1) is established each month from administrative data using the specified algorithm.

Descriptive definitions for being managed for identifiable asthma are as follows. Specifications follow the descriptive definitions. Identifiable asthma is present in any child who has:

- Any prior hospitalization with asthma as primary or secondary diagnosis;

OR

- Other qualifying events, all ages:

o Three or more ambulatory visits with diagnosis of asthma or bronchitis, OR
o Two or more ambulatory visits with a diagnosis of asthma and/or bronchitis AND one or more asthma-related prescriptions

OR

- Other qualifying events, occurring after the fifth birthday:

o One or more prior ambulatory visits with asthma as the primary diagnosis AND a subsequent ED visit in the Reporting Month,
OR
o Two or more ambulatory visits with asthma as a diagnosis,
OR
o One ambulatory visit with asthma as a diagnosis AND at least one asthma-related prescription,
OR
o Two or more ambulatory visits with a diagnosis of bronchitis

For eligibility purposes, asthma-related medicine means long-acting beta-agonist (alone or in combination) or inhaled corticosteroid (alone or in combination), anti- asthmatic combinations, methylxanthines (alone or in combination), and/or mast cell stabilizers. See below further regarding this specification. Note that leukotriene modifiers and short term beta agonists are excluded for the purpose of establishing identifiable asthma. Data from the year prior to the reporting year are used, as well as all months prior to the reporting month in the reporting year (see Appendix Figure 1). Detailed specifications for asthma related medicine can begin with the NCQA NDC list (ASM-C_DASM-C_final_2012, found by clicking through at (http://www.ncqa.org/HEDISQualityMe asurement/HEDISMeasures/HEDIS20 12/HEDIS2012FinalNDCLists.aspx) Eliminate medications in the following 2 categories: leukotriene modifiers, short-acting inhaled beta-2 agonists). May use equivalent updated lists when provided by NCQA. Even if included in NCQA list, we further exclude indacaterol, a recently approved long acting beta agonist that is indicated in the US only for the treatment of COPD.

The analysis should be conducted on a month by month basis as described herein: Within the group of children who meet the criteria for identifiable asthma, identify and maintain a unique patient identifier, age, and all stratification variables. We call the time frame during which eligibility is established to be the Assessment Period.

For each month of the Reporting Year, determine eligibility for each patient, as of the last day of the month prior to the reporting month. This illustration assumes that the Reporting Year is 2011. When assessing January 2011, consider all of Calendar Year 2010 as the Assessment Period for assessing the presence or absences of identifiable asthma. For February, 2011 the Assessment Period includes all of calendar year 2010 AND January 2011. Repeat this progression monthly so that for December, 2011 identifiable asthma one would identify children with identifiable asthma using an Assessment Period from January 2010 through November 2011. For each month, assess whether the continuous enrollment criterion is met prior to including the month in the denominator. For example, for January 2011, the child must have been enrolled in November and

December, 2010 (plus January 2011). Another example, for December 2011, to be eligible the child must have been enrolled in October 2011 and November 2011, as well as December. See Figure 1 in Appendix.

Develop Denominator sample according to Appendix Figure 2 and consistent with the instructions in sections S.18 and S.20.

Codes used for definitions are specified in s2b and include specifications of Hospitalization, Emergency Department Visits, Ambulatory/Office Visits, Asthma diagnosis,

Please note Figures 1 and 2 in the Appendix and the specifications in s2b are considered INTEGRAL to these specifications and are not optional.

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
ED visits that are already in the sample OR Children that fall outside of specified age range of 2-21 OR who do not meet time enrollment criteria OR do not meet identifiable asthma prior to the ED visit, OR children with concurrent or pre-existing COPD, Cystic Fibrosis or Emphysema.

At the discretion of the accountability entity, the denominator may be restricted to children 2-18.

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
Denominator Exclusions
1) Children with concurrent or pre-existing:
a. Chronic Obstructive Pulmonary Disease (COPD) diagnosis;
OR
b. Cystic Fibrosis diagnosis;
OR
c. Emphysema diagnosis.

OR
2) Children without identifiable asthma as specified
OR
3) Outside of specified age range
OR
4) Events occurring in patients who have not been enrolled in the reporting plan for at least two consecutive months before the index reporting month (a total of 3 consecutive months, including the reporting month).

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*
This measure requires stratification by age group. Several additional stratifications are optional but may be required by the accountability entity or provided by the reporting entity. These variables include race/ethnicity, rurality/urbanicity and county level of poverty.

Stratify by age group (reporting entity should specify whether to use age at month of qualifying event or age on first day of reporting year):
•Age 2-5 years (second birthday to the day before the 6th birthday);
•Age 6-11 years (sixth birthday to the day before the 12th birthday);

•Age 12-18 years (twelfth birthday to the day before the 18th birthday); and
•Age 19-21 years (nineteenth birthday to the day before the 21st birthday).

Age strata are to be reported distinctly and not combined.

Optional stratifications require data elements such as:
•Race/Ethnicity
•Insurance type (Public, Commercial, Uninsured)
•Benefit type (if insured): HMO, PPO, Medicaid Primary Care Case Management (PCCM) Plan, Fee for Service (FFS), other
•Zip code, state and county or equivalent area of parent/caregiver's residence. Record FIPS if available

Stratification variables details
•Race/Ethnicity: Hispanic, Non-Hispanic Black, Non-Hispanic White; Non- Hispanic Asian/Pacific Islander, other Non-Hispanic
•Public vs Commercial (Private Insurance).
•HMO vs PPO vs FFS vs PCCM vs other; Within Medicaid, States may ask for reporting of FFS vs Managed Care or other relevant enrollment categories (e.g., TANF, SSI).
•Urban Influence Code. Identify the Urban Influence Code or UIC. (2013 urban influence codes available at: http://www.ers.usda.gov/data-products/urban- influence-codes.aspx#.UZUvG2cVoj8). Use parent or primary caregiver's place of residence to determine UIC. State and county names can be linked or looked up directly or zip codes can be linked to county indirectly, using the Missouri Census Data Center (http://mcdc.missouri.edu/). These data will link to county or county equivalents as used in various states.

Urban Influence Codes (UIC) have been developed by the USDA to describe levels of urbanicity and rurality. While each UIC has its own meaningful definition, some researchers choose to aggregate various codes. Well regarded schemas for aggregation of codes include Bennett and colleagues at the South Carolina Rural Research Center. Their aggregation scheme brings together Codes 1 & 2 as Urban; 3,5, & 8 as micropolitan rural; 4,6, & 7 as rural adjacent to a metro area; and 9, 10, 11, & 12 as remote rural. We acknowledge that UIC 5 (adjacent rural area) may appropriately be aggregated with 4,6,&7 as rural. Frontier health care may be approximated by analysis of the remote rural categories (UIC 9, 11 and 12). Alternatively, Gary Hart, Director of the Center for Rural Health at the University of North Dakota School of Medicine & Health Science suggests that UIC 9-12 is the best overall approach to using county level data to study frontier health. Inclusion of UIC 8 would make the analysis more sensitive to including frontier areas but at a meaningful cost in specificity.

Those interested in care specific to large cities may wish to aggregate the rural area and analyze UIC 1 and 2 separately.

When stratifying by urbanicity or UIC, the reporting and accountability entities should specify clearly what if any aggregating schema was used.

•Identify the Level of Poverty in the parent or primary caregiver's county of residence. The percent of all residents in poverty by county or county equivalent are available from the US Department of Agriculture at http://www.ers.usda.gov/data-products/county-level-data-sets/download- data.aspx. Our stratification standards are based on 2011 US population data that we have analyzed with SAS 9.3. Using parent or primary caregiver's state and county of residence (or equivalent) or FIPS code, use the variable PCTPOVALL_2011 to categorize into one of 5 Strata:
o Lowest Quartile of Poverty if percent in poverty is <=12.5%
o Second Quartile of Poverty if percent in poverty is >12.5% and <=16.5%
o Third Quartile of poverty if percent in poverty is >16.5% and <=20.7%
o First Upper Quartile (75th-90th) if percent in poverty is >20.7% and <=25.7%
o Second Upper Quartile (>90th percentile)

These classification standards may be updated by the accountability entity suing more recent data if desired.

Note: if needed, the Missouri Census Data Center may be used to link zip codes to county equivalents. http://mcdc.missouri.edu/

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
Stratification by risk category/subgroup
If other:

S.12. **Type of score:**
Rate/proportion
If other:

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

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
Step 1: Select starting cohort
Identify the upper age limit to be used, either 18 or 21. The measure is specified from 2 to 21 years, with 19-21 year olds considered optional at the discretion of the accountability entity.

Appendix Figures 1 and 2 and Data Codes in S2b provide an overview and guide for eligibility and sample selection.

Step 2: Conduct analysis of administrative data using the specifications described in denominator description to identify children within the specified age range with identifiable asthma. The analysis should be conducted on a month by month basis as described herein:

Determine eligibility for each patient, as of the last day of the month prior to the reporting month. For example, if the goal is to report for January 2011, first identify children with identifiable asthma (above), and analyze all of calendar year 2010 when doing so. Continuous enrollment criterion requires that the child was enrolled in November and December of 2010. Next, for February analyze all of calendar year 2010 AND January 2011. Continuous enrollment criterion requires that the child was enrolled in December 2010 and January 2011. Repeat this progression monthly so that for December, one would identify children with identifiable asthma and analyze all of calendar year 2010 AND January through November 2011 when doing so. Continuous enrollment criterion requires that for December the child was also enrolled in October 2011 and November 2011. Appendix Figure 1 describes and illustrates the month by month analysis.

Step 3: Identify ED Visits and hospitalizations for asthma in eligible children.
Considering only the children who were identified as eligible in the given month
according to Step 2, perform a month-by-month analysis to identify and log all ED visits with asthma as a primary or secondary diagnosis and all hospitalizations with asthma as a primary or secondary diagnosis for each reporting month, using specifications described in denominator and the codes described above and in S2b. Maintain stratification data elements, age, and unique identifiers.

Step 4: Stratify by age and develop random samples.
Stratify by age group (use age at month of qualifying event):

- Age 2-5 years (second birthday to the day before the 6th birthday);
- Age 6-11 years (sixth birthday to the day before the 12th birthday);
- Age 12-18 years (twelfth birthday to the day before the 18th birthday); and
- Age 19-21 years (nineteenth birthday to the day before the 21st birthday).

For each age group develop a random sample of 500 events as described in the sampling section below and illustrated in Appendix Figure 2.

Appendix Figure 2 is necessary to guide sample development. Several key remarks may help Figure 2 to be more understandable:

Before sample selection can be randomized, eligibility needs to be determined based on 3 key factors:
- Identifiable asthma diagnosis AND
- Month by month time analysis AND
- Asthma emergency department (ED) visit OR Asthma hospitalization

After eligibility is determined, the randomized sample can fall into one of three groups only:
A. Asthma ED visit only OR
B. Asthma hospitalization on same day as ED visit OR
C. Asthma hospitalization only

A. Asthma ED visit only qualifies for (at least) denominator inclusion
B. Asthma hospitalization on same day as ED visit qualifies for denominator AND numerator inclusion
C. Asthma hospitalization only needs further investigation to determine denominator inclusion
. • Do NOT include in denominator
-- if sample was not hospitalized (admitted) from an asthma ED Visit, OR
-- if ED visit was already in the sample under any criteria (avoid duplication)
ELSE,
. • Do include in both Denominator AND Numerator

Step 5: Collect stratification data elements from administrative data.
Collect the following data elements for all eligible children in each randomized sample. These data elements are used for reporting stratified results. Entities that are interested in assuring large samples for specific stratified analyses may choose to incorporate a further stratified sampling scheme and oversample to assure that there is a sample size of 100-500 per stratification category (e.g. race or ethnicity of interest). Such a sampling scheme must employ an appropriate weighting system (using the reciprocal of the likelihood for selection as a weight, c.f. Rao, P., 2000. Sampling Methodologies with Applications. New York: Chapman & Hall) to estimate overall performance. Alternatively, the stratified samples may be used only for reporting stratum specific performance comparison and not for estimating the overall performance. Approximate 95% confidence interval widths (assuming a rate of 50% appropriateness) are shown in the sampling specifications. We specify to oversample by 25% to account for potential loss in our event identifications.

Stratification data elements include:
- Race
- Ethnicity
- Insurance type (Public, Commercial, Uninsured)
- Benefit type (if insured): HMO, PPO, Medicaid Primary Care Case Management (PCCM) Plan, Fee for Service (FFS), other
- Zip code, state and county or equivalent area of parent/caregiver's residence. Record FIPS if available

Step 6: Categorize stratification variables as described in the stratification section S.12.

Step 7. Conduct Chart Audit (Medical Record Review) of GROUP A ED Visits.
Group A ED visits that have been selected for inclusion in the sample require a chart audit to assess eligibility for the numerator based on the explicit appropriateness criteria. They have already qualified for inclusion in the denominator. Eligibility for the numerator is established based on documentation of any of the following items. Review may be terminated once any qualification for the numerator is identified.

- Disposition of the child from the ED was admission to an inpatient hospital, OR
- Documented physical findings consistent with respiratory distress, including:
  - o Labored breathing with retractions and/or grunting; or
  - o Labored breathing with evidence of accessory muscle use; or,
  - o Markedly decreased breath sounds; OR
- O2 saturation level below 90% documented in the ED; OR
- An ABG obtained and reported in the ED; OR
- The child had a consultation with a pulmonologist or asthma specialist that was ordered and provided in the ED; OR
- Specific documentation that:
  - o The child was referred to the ED after evaluation by the PCP or other licensed clinician practitioner; OR
  - o The child received two or more doses of inhaled rescue medications without sufficient clinical improvement; OR
  - o The child was assessed with an objective instrument such as a peak flow meter and was found to be in a pre-defined “red zone” of peak flow measurement as part of a pre-specified asthma action or similar plan.

There is no specified order for review. Some institutions may prefer to record all reasons for numerator qualification to support ongoing or planned improvement activities.

Note 1: Evidence for hospitalization above requires that the child was admitted to any hospital as an inpatient. This includes admission directly to a medical or pediatric ICU or inpatient floor or transfer directly to an inpatient facility. If a child is transferred to another hospital, confirmation that the child actually was admitted directly (i.e., was not first admitted to another ED prior to admission) is necessary prior to qualifying for the numerator. Such confirmation may include evidence from the administrative data review in Step 2. Other potential sources for this information include ED discharge summary, disposition on a flow, admit, or discharge form, or documentation by doctors, nurses, nurse practitioners or physician assistants.

Note 2: Evidence that the child was referred to the ED requires documentation of both of two requirements. The requirements are:

- The child/adolescent was referred by a clinician to come to the ED; and
- The child/adolescent was evaluated by the clinician prior to referral. Generally such evaluations will be in person. Assessment of respiratory distress by listening or speaking to the child/adolescent over the telephone is sufficient if such an examination is clearly documented. Report of this requirement being met by the child/adolescent or parent/caregiver is sufficient to meet this criterion. Report of contact from the referring physician can also fulfill this criterion. Nursing notes, triage notes and clinician notes, particularly history of present illness (HPI) are common sources for this data.

Note 3: Evidence of a parent or caregiver report that the child received two or more doses of an inhaled rescue medication with insufficient clinical improvement typically will be found in triage, nursing, clinician, or respiratory therapy notes. It may also be documented as a part of medication reconciliation during intake. It requires documentation:

- That multiple treatments of medication were provided by inhalation or injection prior to arrival in the ED;
- That the medication(s) provided were specifically rescue medications and are not a part of the of the child/adolescent’s preventive or maintenance regimen; and,

• That the child continued to be in distress following the treatments (alternately that the child did not improve substantially).

Note 4: Parent / caregiver report that their child was in a pre-defined "red zone" of peak flow measurement includes documentation:

• That a pre-specified asthma plan (action plan) exists and defines a "red zone" based upon an objective respiratory measurement, such as a peak flow rate; and

• That the objective assessment was made prior to coming to the ED and that the results were in the pre-specified "red zone."

Note 5: Reports of the physical exam typically may be found on triage, nursing, physician, nurse practitioner, physician assistant, or respiratory therapist notes. Diverse language may be used to describe similar findings, for example:

• The term pulling may be used to describe retractions. Retractions may be described as nasal flaring (particularly in infants), or by location (see below);

• Increased work of breathing may be indicated or it may be described by physical findings such as the use of accessory muscles, such as sub or intercostal muscles, supraclavicular or suprasternal. "Mildly" increased work of breathing or "minimal" retractions do not meet these criteria.

• Labored breathing, significant increased work of breathing, respiratory distress (moderate or greater), difficulty breathing, poor air entry (or air exchange or air movement) may all describe findings that meet this criterion. Grunting indicates that the child or adolescent is generating clearly audible sounds with each breath concomitant with apparent increased work of breathing. These may be found in the general description or respiratory section of the physical exam.

• Markedly (or severely) reduced breath sounds and descriptions of poor air movement are typically a part of an auscultation during the pulmonary exam.

Note 6: Documented evidence of the percent oxygen (O2) saturation from a transcutaneous assessment can be located in a flow sheet, nursing, respiratory therapy, or physician/nurse practitioner/physical assistant note or may be recorded as part of the physical exam. The O2 saturation may be obtained initially at triage and is often assessed periodically during the visit. Any O2 saturation less than 90 satisfies the criteria.

Note 7: An ABG requires drawing of a blood specimen from an artery and is distinguished from a venous blood gas, which would not fulfill this criterion. This typically would be found in a laboratory results section of the record or commented as a finding in a clinician's note, such as a respiratory therapist, doctor, PA, NP, or RN. An ABG is typically comprised of at least a pO2, pCO2, and pH.

Note 8: Consultation with a pulmonary specialist or other asthma specialist requires both an order for such a physician consultation and evidence that the consultation occurred, including a note from the consultant specialist. Typically a consultation from a pulmonologist, pediatric pulmonologist, allergist, or pediatric allergist would fill this criterion.

Identify which ED visits meet at least one criterion for the Numerator.
Maintain stratification variables.

Step 8: Conduct Chart Audit (Medical Record Review) to Assess Eligibility of GROUP C Hospitalizations for Inclusion in Denominator.
Within each stratification group (as determined above), identify the asthma hospitalizations for which there were not associated ED visits (Group C) found in the administrative data. An asthma ED visit and asthma hospitalization are said to be associated on the basis of the administrative data review only if they occur on the same service data and at the same institutions and if the hospital discharge date is after the ED service date. Such hospitalizations should have been included in Group B. Other hospitalizations require a review of the

medical record to determine if they were admitted or transferred directly from an ED visit that was not otherwise in the sample (i.e., was not identified via the administrative data analysis).

The chart audit/medical record review seeks evidence that the child was admitted to the hospital directly from the ED or transferred directly from another hospital's ED. Evidence may include an ED note (physician, nurse, physician assistant, nurse practitioner), flow, or face sheet that indicates the disposition of the ED visit was hospital admission.

It may also include a note from within the hospitalization (including the admission note or any physician, nurse, physician assistant, nurse practitioner note), flow sheet, face sheet, or discharge summary that indicates that the hospitalization came directly from (was admitted from or transferred directly from) an ED. In either case, the ED visit is only eligible for inclusion if the chart review specifies the date and institution of the ED visit sufficiently to assure that it can be uniquely identified and all duplication avoided. Others are excluded.

For example if an ED visit was identified in Group A and the resulting hospitalization appeared in Group C (either because of a different service date or different institution), the Group A ED visit would be included and the Group C hospitalization excluded as a duplicate (even though there was a preceding ED visit). If the child is uniquely included in the sample for that month and there is clear evidence that the admission came directly from an ED (e.g., was not transferred from another hospital after having been admitted from the ED) this measure can be satisfied. De-duplication requires the elimination of any duplications that remain in the sample, considering the unit of analysis to be the ED visit. In other words, all ED visits must be included only once. Further, an ED visit identified via the hospitalization that also was a transfer from another ED visit already in the sample should have been removed as a duplicate. Similarly all hospitalizations lacking sufficient document that the child was admitted or transferred directly from an ED visit or lacking sufficient detail to allow confirmation that the ED visit referred to in the notes is not already in the sample elsewhere (e.g., from Group A) should have been removed.

Those Group C hospitalizations that can be identified as resulting from a unique (unduplicated) ED visit are included in BOTH the numerator and the denominator.

Step 9: Calculate and report the measure.
a) For each age stratum, count the number of events in the sample that qualify for the denominator (ND).
b) For each age stratum, count the number of events in the sample and in the denominator that qualify for the numerator (NN).
c) For each stratum, calculate the percent of appropriate ED visits as Percent Appropriate = 100 * (NN / ND). Report to one decimal place.

Step 10: Report each stratification category listed below, that have an N of at least 50.
a) Race and ethnicity
b) Insurance type (Public/Medicaid, Private/Commercial, None, other)
c) Benefit type: HMO vs PPO vs FFS vs PCCM vs other
d) Urban Influence Code or UIC.
e) Level of poverty in the county of residence.

Step 11. Calculate and report 95% confidence intervals (using binomial distribution for each stratum) for each age specific stratum and for all of the Step 9 stratifications.

a) Calculate the standard error as the square root of each proportion by [1-the same proportion] divided by the number in the denominator.
b) Multiply the standard error by 1.96.
c) Subtract that value from the measured proportion. Report the greater of 0 and that number as the lower bound of the 95% confidence interval.

d) Add the product from b to the measured proportion. Use the lesser of that sum or 1 as the upper bound of the 95% confidence interval.
e) To report as percent, multiply by 100.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Children and adolescents who have a qualifying ED visit associated with asthma as the first or second diagnosis and are in the random sample;

AND have at least one of the following:

•Disposition of the ED visit was admission to the hospital,
OR
•Documented physical findings consistent with respiratory distress, including any of the following:
o Labored breathing (including moderate or severe increased work of breathing); OR
o Retractions, grunting, and/or evidence of accessory muscle use; OR
o Markedly decreased breath sounds;
OR
•Recorded oxygen saturation below 90%;
OR
•An arterial blood gas (ABG) was obtained in the emergency department;
OR
•The child had a consultation with a pulmonologist or asthma specialist that was ordered and provided in the ED;
OR
•There is clear documentation that prior to arrival in the ED any of the following occurred:
o The child was referred to the ED after evaluation by the PCP or other clinician. The evaluation may include an in person visit or auscultation including via telephone. OR
o The child received two or more doses of inhaled rescue medications without sufficient clinical improvement. Documentation of parent report meets the criterion. OR
o The child was assessed with an objective instrument such as a peak flow meter and was found to be in a pre-defined "red zone" of peak flow measurement as part of an asthma action or similar plan. Documentation requires ALL of the following: a written asthma action plan exists AND defines a "red zone" for which urgent assessment by a clinician is indicated AND an objective assessment was made and its result was in the pre-defined red zone. Documentation of parent report meets the criterion.

The denominator is a random sample of the patients in each age stratum who have visited the emergency department for asthma (as a first or second diagnosis) and meet the specified criteria for having identifiable asthma (defined in s2b).

Separate numerators and denominators are reported for children age 2-5, 6-11, 12-18, and, optionally, 19-21 years. An overall rate across strata is not reported.

Denominator Elements:

The presence of identifiable asthma (see table 1) is established each month from administrative data using the specified algorithm.

Descriptive definitions for being managed for identifiable asthma are as follows. Specifications follow the descriptive definitions. Identifiable asthma is present in any child who has:

- Any prior hospitalization with asthma as primary or secondary diagnosis;

OR

- Other qualifying events, all ages:

o Three or more ambulatory visits with diagnosis of asthma or bronchitis, OR
o Two or more ambulatory visits with a diagnosis of asthma and/or bronchitis AND one or more asthma-related prescriptions

OR

- Other qualifying events, occurring after the fifth birthday:

o One or more prior ambulatory visits with asthma as the primary diagnosis AND a subsequent ED visit in the Reporting Month,
OR
o Two or more ambulatory visits with asthma as a diagnosis,
OR
o One ambulatory visit with asthma as a diagnosis AND at least one asthma-related prescription,
OR
o Two or more ambulatory visits with a diagnosis of bronchitis

For eligibility purposes, asthma-related medicine means long-acting beta-agonist (alone or in combination) or inhaled corticosteroid (alone or in combination), anti- asthmatic combinations, methylxanthines (alone or in combination), and/or mast cell stabilizers. See below further regarding this specification. Note that leukotriene modifiers and short term beta agonists are excluded for the purpose of establishing identifiable asthma. Data from the year prior to the reporting year are used, as well as all months prior to the reporting month in the reporting year (see Appendix Figure 1). Detailed specifications for asthma related medicine can begin with the NCQA NDC list (ASM-C_DASM-C_final_2012, found by clicking through at (http://www.ncqa.org/HEDISQualityMe asurement/HEDISMeasures/HEDIS20 12/HEDIS2012FinalNDCLists.aspx) Eliminate medications in the following 2 categories: leukotriene modifiers, short-acting inhaled beta-2 agonists). May use equivalent updated lists when provided by NCQA. Even if included in NCQA list, we further exclude indacaterol, a recently approved long acting beta agonist that is indicated in the US only for the treatment of COPD.

The analysis should be conducted on a month by month basis as described herein: Within the group of children who meet the criteria for identifiable asthma, identify and maintain a unique patient identifier, age, and all stratification variables. We call the time frame during which eligibility is established to be the Assessment Period.

For each month of the Reporting Year, determine eligibility for each patient, as of the last day of the month prior to the reporting month. This illustration assumes that the Reporting Year is 2011. When assessing January 2011, consider all of Calendar Year 2010 as the Assessment Period for assessing the presence or absences of identifiable asthma. For February, 2011 the Assessment Period includes all of calendar year 2010 AND January 2011. Repeat this progression monthly so that for December, 2011 identifiable asthma one would identify children with identifiable asthma using an Assessment Period from January 2010 through November 2011. For each month, assess whether the continuous enrollment criterion is met prior to including the month in the denominator. For example, for January 2011, the child must have been enrolled in November and December, 2010 (plus January 2011). Another example, for December 2011, to be eligible the child must have been enrolled in October 2011 and November 2011, as well as December. See Figure 1 in Appendix.

Develop Denominator sample according to Appendix Figure 2 and consistent with the instructions in sections S.18 and S.20.

Codes used for definitions are specified in s2b and include specifications of Hospitalization, Emergency Department Visits, Ambulatory/Office Visits, Asthma diagnosis,

Please note Figures 1 and 2 in the Appendix and the specifications in s2b are considered INTEGRAL to these specifications and are not optional.

ED visits that are already in the sample OR Children that fall outside of specified age range of 2-21 OR who do not meet time enrollment criteria OR do not meet identifiable asthma prior to the ED visit, OR children with concurrent or pre-existing COPD, Cystic Fibrosis or Emphysema.
At the discretion of the accountability entity, the denominator may be restricted to children 2-18.

Denominator Exclusions
1) Children with concurrent or pre-existing:
a. Chronic Obstructive Pulmonary Disease (COPD) diagnosis;
OR
b. Cystic Fibrosis diagnosis;
OR
c. Emphysema diagnosis.

OR
2) Children without identifiable asthma as specified
OR
3) Outside of specified age range
OR
4) Events occurring in patients who have not been enrolled in the reporting plan for at least two consecutive months before the index reporting month (a total of 3 consecutive months, including the reporting month).

This measure requires stratification by age group. Several additional stratifications are optional but may be required by the accountability entity or provided by the reporting entity. These variables include race/ethnicity, rurality/urbanicity and county level of poverty.

Stratify by age group (reporting entity should specify whether to use age at month of qualifying event or age on first day of reporting year):
•Age 2-5 years (second birthday to the day before the 6th birthday);
•Age 6-11 years (sixth birthday to the day before the 12th birthday);
•Age 12-18 years (twelfth birthday to the day before the 18th birthday); and
•Age 19-21 years (nineteenth birthday to the day before the 21st birthday).

Age strata are to be reported distinctly and not combined.

Optional stratifications require data elements such as:
•Race/Ethnicity
•Insurance type (Public, Commercial, Uninsured)
•Benefit type (if insured): HMO, PPO, Medicaid Primary Care Case Management (PCCM) Plan, Fee for Service (FFS), other
•Zip code, state and county or equivalent area of parent/caregiver's residence. Record FIPS if available

Stratification variables details
•Race/Ethnicity: Hispanic, Non-Hispanic Black, Non-Hispanic White; Non- Hispanic Asian/Pacific Islander, other Non-Hispanic
•Public vs Commercial (Private Insurance).
•HMO vs PPO vs FFS vs PCCM vs other; Within Medicaid, States may ask for reporting of FFS vs Managed Care or other relevant enrollment categories (e.g., TANF, SSI).

•Urban Influence Code. Identify the Urban Influence Code or UIC. (2013 urban influence codes available at: http://www.ers.usda.gov/data-products/urban- influence-codes.aspx#.UZUvG2cVoj8). Use parent or primary caregiver's place of residence to determine UIC. State and county names can be linked or looked up directly or zip codes can be linked to county indirectly, using the Missouri Census Data Center (http://mcdc.missouri.edu/). These data will link to county or county equivalents as used in various states.

Urban Influence Codes (UIC) have been developed by the USDA to describe levels of urbanicity and rurality. While each UIC has its own meaningful definition, some researchers choose to aggregate various codes. Well regarded schemas for aggregation of codes include Bennett and colleagues at the South Carolina Rural Research Center. Their aggregation scheme brings together Codes 1 & 2 as Urban; 3,5, & 8 as micropolitan rural; 4,6, & 7 as rural adjacent to a metro area; and 9, 10, 11, & 12 as remote rural. We acknowledge that UIC 5 (adjacent rural area) may appropriately be aggregated with 4,6,&7 as rural. Frontier health care may be approximated by analysis of the remote rural categories (UIC 9, 11 and 12). Alternatively, Gary Hart, Director of the Center for Rural Health at the University of North Dakota School of Medicine & Health Science suggests that UIC 9-12 is the best overall approach to using county level data to study frontier health. Inclusion of UIC 8 would make the analysis more sensitive to including frontier areas but at a meaningful cost in specificity.

Those interested in care specific to large cities may wish to aggregate the rural area and analyze UIC 1 and 2 separately.

When stratifying by urbanicity or UIC, the reporting and accountability entities should specify clearly what if any aggregating schema was used.

•Identify the Level of Poverty in the parent or primary caregiver's county of residence. The percent of all residents in poverty by county or county equivalent are available from the US Department of Agriculture at http://www.ers.usda.gov/data-products/county-level-data-sets/download- data.aspx. Our stratification standards are based on 2011 US population data that we have analyzed with SAS 9.3. Using parent or primary caregiver's state and county of residence (or equivalent) or FIPS code, use the variable PCTPOVALL_2011 to categorize into one of 5 Strata:
o Lowest Quartile of Poverty if percent in poverty is <=12.5%
o Second Quartile of Poverty if percent in poverty is >12.5% and <=16.5%
o Third Quartile of poverty if percent in poverty is >16.5% and <=20.7%
o First Upper Quartile (75th-90th) if percent in poverty is >20.7% and <=25.7%
o Second Upper Quartile (>90th percentile)

These classification standards may be updated by the accountability entity suing more recent data if desired. Note: if needed, the Missouri Census Data Center may be used to link zip codes to county equivalents. http://mcdc.missouri.edu/

Within each age group, randomly select 500 ED visits among those identified in Step 4. Analyze each age strata's random sample distinctly:

Sort into three groups according to Appendix Figure A.1.b.

• Group A: Those with asthma ED visits ONLY and no associated asthma hospitalization to the same hospital on the same date. These ED visits are INCLUDED in the Denominator and receive Medical Record Review to assess eligibility for the Numerator;

• Group B: Those with both asthma ED Visits and asthma Hospitalizations at the same facility on the same date and for whom the hospital discharge date is after the ED date of service. These ED visits are INCLUDED in both the Denominator and in the Numerator. No further review is necessary to establish appropriateness;

• Group C: Those with asthma Hospitalizations ONLY and no associated asthma ED Visit to the same hospital on the same date. Please note that children admitted to the ED one date and admitted to the hospital the next day (from the same ED visit) will be identified in this group. Group C Hospitalizations are subject to Medical Record Review to assess eligibility for the Denominator. If they are eligible for the denominator they will be included in BOTH the Numerator and Denominator.

Please note that the terms medical chart and medical record are used interchangeably, as are the terms audit and review in this context.
Figure 2 in the Appendix is integral to sample selection and no optional.
Notes:
•Determining eligibility for sample selection precedes determining eligibility for measure.
•On the basis of the Administrative Data Analysis, children who are potentially eligible for the measure will be identified and segregated into Groups A, B, and C (the blue boxes In Figure 2 of Appendix).
•Children are eligible for Group B if three things are found in the administrative data: ED Visit; Hospitalization on same day and same institution; and Hospital discharge is after date of ED visit.
•National and NY State data suggest that approximately ¾ of childhood asthma hospitalizations are admitted from ED, that about 1 in 9 childhood asthma ED visits result in hospitalization and that children admitted from the ED may not have their ED visit coded in administrative data.
•Medical record review determines eligibility for numerator among the Group A children, all of whom have already qualified to be included in the denominator.
•Group B children are eligible for both the numerator and the denominator on the basis of administrative data analysis alone and do not require chart review.
•Medical record review determines eligibility for inclusion in the measure (denominator!) for Group C children. If they are eligible for the denominator (i.e. that have been admitted directly from an unduplicated ED visit) then they are also qualified for the numerator.

The impact of sample size on the width of the confidence interval is illustrated by assuming 50% appropriateness and a variety of sample size to calculate the width of the confidence intervals around the estimate obtained above. Variations from 50% will bring down the size of the confidence interval.

N=	50,	+ / - 13%
N=	75,	+ / - 11%
N=	100,	+ / - 10%
N=	150,	+ / - 8%
N=	200,	+ / - 7%
N=	250,	+ / - 6%
N=	400,	+ / - 5%

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Claims (Only), EHRs Hybrid, Paper Records

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

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

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Emergency Department, Hospital
If other:

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

2. Validity – See attached Measure Testing Submission Form
nqf_testing_attachment_12_20_16_LKV.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.3 For maintenance of endorsement
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

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

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

What were the statistical results of the analyses used to select risk factors?

Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)
No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma

Date of Submission: 12/20/2016

Type of Measure:

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

Instructions

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

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

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

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

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):
- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR
- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

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

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

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.

$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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

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

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

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

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

New York State Medicaid claims data 2010 – 2012.

Our work builds off of work performed by our CAPQuaM partner and steering committee member, NCQA. For specific data reliability and signal to noise analyses, we incorporate by reference (and will present more selectively) NCQA data relevant to their submission for NQF – endorsed asthma related measures 0036, 1799, and 1800.

Their analyses demonstrate the capacity to use administrative data to identify the applicable denominator population. There is nearly complete overlap of the denominator codes and there is overlap of the denominator elements. Where codes differ it is specific to decisions made by the CAPQuaM expert panel which was aware of the NCQA measures.

Newly abstracted data was also used for this measure.

1.3. What are the dates of the data used in testing? 10/2009 – 11/2013

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

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☒ health plan	☒ health plan
☐ other: Integrated delivery system, population, state, region, county	☐ other: Integrated delivery system, population, state, region, county

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

We surveyed 9 hospitals around the country for data availability, data source, and ease of data abstraction related to this measure. Most of these hospitals used electronic medical records, but we also surveyed several that at the time did not. Respondents included: VP Quality & Patient Safety; Performance Improvement Coordinator; Director of Medical Affairs, QI RN, Quality and Performance Improvement Project Management Coordinator; Clinical Manager Emergency Services & the Resource Pool; Case Manager Women and Children's Service Line; and the Director of Quality Management. 8 of the 9 specifically indicated that they were answering with an institutional perspective and 5 of the nine specifically reviewed charts to assist in answering the survey. At least 19 charts were reviewed.

All 9 respondents indicated that race/ethnicity were not difficult to collect and was present in the medical records. The answer set regarding difficulty of abstraction included: Not Available; Not Difficult to Collect; Difficult to Collect; and Very Difficult to Collect. Eight of nine indicated that date of birth was not difficult to collect and all agreed that it was in the medical record, to allow for simple assessment of age. All 9 agreed that payment source was likewise not difficult to collect from the patient's record. All indicated that the presence of prior asthma was typically in the record and not difficult to assess via chart audit. Eight of nine indicated that abstraction of clinical data to assess the appropriateness of the ED was not difficult to accomplish via chart audit. In contrast 2 of 9 indicated that a forma asthma severity score assessment was not available in the ED chart, and three of nine that no formal interpretation of such a score would be available in the chart.

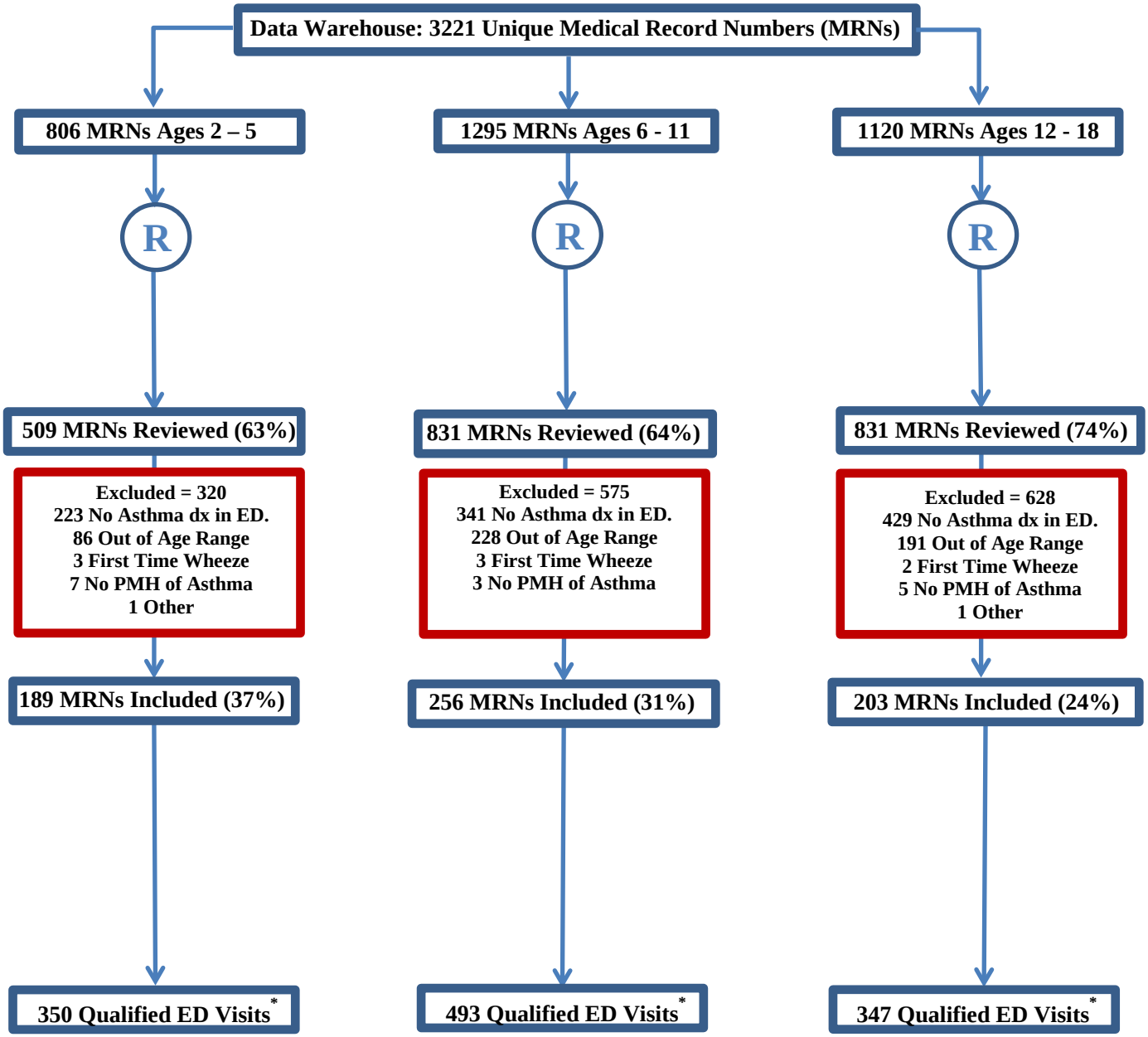
Nine of nine indicated that the following were not difficult to collect: collection of oxygen saturation; identifying the lowest recorded oxygen saturation; identifying whether or not an arterial blood gas (ABG) had been collected was not difficult to collect; identifying the level of respiratory distress (mild, moderate, severe dyspnea); presence of retractions; and admission to the hospital from the ED. More varied was information regarding whether or not the child had been referred into the ED by the PCP (4 = "Not Difficult", 2= "Difficult"; 2= "Very Difficult", 1="Very Difficult", and 2="Not Available"). This last finding was among those aspects of testing that led our specifications to name indicate appropriateness based on evidence of presence of at least one of the criteria, and not to consider the absence to be meaningful in and of itself. We further extend this principle to recognize that certain

aspects of care that are based upon parental response are not likely to fully captured in the medical record, even if such documentation ought to be part of a standard of high quality care. Thus we speak of “Appropriate” and “Questionable” ED visits, rather than “inappropriate” visits.

Foundational analyses for this measure included:

Analysis of NY State Medicaid Managed Care claims data, including claims from all MCO’s that are contracted for Medicaid care by our partner, the NY State Dept of Health. We identified eligible populations and events from both RY 2011 and 2012 and include children from counties in nine urban influence codes and in counties poverty level 1-3. NY State does not have any counties in the lowest 25% of poverty or with UIC of 10-12. New York has more than 60 counties and numerous health plan vendors. Analysis in year 2011 provided very similar data to 2012

Figure 1: Chart Selection for Review of ED Appropriate Use for Pediatric Asthma



*** Some patients had multiple ED visits for asthma**
R = Randomized

.

For the NCQA analysis, nine health plans covering a variety of geographic areas within the United States were asked to provide a complete administrative data file consisting of any member in their commercial and Medicaid product lines for anyone that had a diagnosis code for asthma during the calendar years of 2009-2010. The complete member-level administrative file used for analysis included a total of more than 82,000 health plan members with asthma.

The specific measure demonstration and testing was done at one site, a New York City Academic Medical Center. In this testing, **sample selection can be summarized in the diagram above**.

The eligible observation period was October 2009 to November 2013. Please note, because of the limitations of the data systems available for testing randomization happened at the level of the patient. For patients with 1-3 visits for asthma in the included time frame, we included all visits. For patients with more than 3 visits, the first three visits were included. Hence the average number of visits per child was 350/189= 1.9 for the younger children, 493/256=1.9 for the school age children, and 347/203=1.7 for adolescents. In NY State Medicaid in 2011, the median number of visits per child was 1, the 75th percentile was 2 and the 90th percentile was 3 (N=26,169 children). Hence this finding is plausible and consistent, given the 4 year time frame that we sampled.

The eligible observation period was October 2009 to November 2013. Please note, because of the limitations of the data systems available for testing randomization happened at the level of the patient. For patients with 1-3 visits for asthma in the included time frame, we included all visits. For patients with more than 3 visits, the first three visits were included. Hence the average number of visits per child was 350/189= 1.9 for the younger children, 493/256=1.9 for the school age children, and 347/203=1.7 for adolescents. In NY State Medicaid in 2011, 40,855 children experienced 61,327 eligible emergency department visits. The median number of visits was 1.5. 10% of children had 3 or more eligible visits and the mode number of visits per child was 1. Hence the findings in our testing is plausible and consistent, given the 4 year time frame that we sampled.

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

Refer to Figure 1 in 1.5. Using the institutional data warehouse, we randomly identified medical record numbers of children who had both ED visits and asthma diagnoses in the specified time frame. Because we were not using claims data to select them, charts had to be reviewed for evidence of prior asthma and to assure that ED visit and asthma diagnosis were concurrent and in the selection time frame. ED visits were excluded if there was not evidence that they were known to be asthmatic, if the ED visit did not have asthma as the first or second diagnosis, or if the ED visit was not in the specified time frame. We included up to 3 visits per selected child, using the first 3 visits when more than three were present. Inclusion criteria included an ED visit with previously established asthma as a primary or secondary diagnosis as documented in the electronic medical record. We developed 3 samples stratified by age: 2-5 years, 6-11 years, and 12-18 years. For ages 2 – 5, we included 350 visits; ages 6 – 11, 493 visits; ages 12 – 18, 347 visits. So included in the measure testing was a total of 1200 ED visits were included in the chart review testing.

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

Assessment of the capacity to identify the eligible population and qualifying events was performed in NY State Medicaid data in both 2011 and 2012 reporting years.

Our construct for the CAPQuaM measure was defined by the multidisciplinary national expert panel using a RAND type modified Delphi process. The panel initially used the term persistent asthma to

describe asthma that was pre-existing and should have been recognized as asthma by the health care system prior to the timing of the ED visit. This construct was renamed by our stakeholder group to be identifiable asthma to avoid confusion with other uses of the term persistent asthma. The construct was intended to be more inclusive than HEDIS' persistent asthma diagnosis, while still removing from consideration those whose asthma was unlikely to have been actively managed at the time.

Holding steady the continuous enrollment criterion at 12 months, HEDIS criteria identified a rate of persistent asthma of 3.1% with the CAPQuaM criteria identifying identifiable asthma at a rate of 8.6%. This ratio is 2.8, which is between 2-3, which is what we had predicted (based on the team's reading of the literature) and was the goals we were hoping to achieve with our criteria and was interpreted to suggest construct validity for our measure. Using data form the National Survey of Children's Health, we estimated the expected rate of asthma in the NY State Medicaid child population to be between 15 - 16%, indicating that our criteria did provide a meaningful filter as we had intended.

We found that by reducing the continuous enrollment period down to three months as was suggested by members of our steering committee that we could increases the number of children eligible for the measure by several tens of thousands while still restricting the measure to those who had received sufficient care for asthma to be identified, and requiring continuous enrollment for attribution to the extent felt important by our multi-stakeholder group.

Assessment of data elements for identifying a population with asthma was performed by NCQA in nine geographically diverse managed care plans.

Assessment of appropriateness was performed in 1200 pediatric ED visits from a single medical center.

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

Insurance status and race/ethnicity for the single site analysis.

Race, ethnicity, zip code, level of poverty in the zip code of caregiver residence, and urban influence in the county of caregiver residence for the NY State analysis.

2a2. RELIABILITY TESTING

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

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

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

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

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

Validity testing was performed at the data element level for both the numerator and the

denominator.

See section 2b2 for validity testing of data elements

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

As noted below, reliability for finding the presence or absence of appropriateness criteria could be trained and kappas suggest high reliability.

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

2b2. VALIDITY TESTING

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

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

☒ **Performance measure score**

☐ **Empirical validity testing**

☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

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

Please see descriptions of testing above as well.

As described below, the literature also supports the use of claims data to identify the presence of asthma. The table summarizes these findings.

Data element	**Reference** (e.g., Quam, et al., 1993)	**Data source** (e.g., Medicare FFS outpatient data)	**Statistical results** (e.g., kappa, sensitivity, specificity, etc.)
Numerator: Validation of Numerator Data Elements was performed by the CAPQuaM development team and the results are summarized in this section following the table.			
Denominator			
Age	NYSDOH CAPQuaM Analysis – internal testing	NY State Medicaid Data	Meaningful variation by age groups as predicted, with peaks in younger children and older adolescents.
	CMS MMIS data requirements Exemplar specifications at https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MSIS/downloads/msisdd2010.pdf	State Medicaid MMIS systems	States are required to submit validated claims data including age or date of birth with a tolerance of 0.1%
Asthma diagnosis in inpatient/ED setting	Wilchesky, M., Tamblyna, R. M., & Huang, A. (2004). Validation of diagnostic codes within medical services claims. Journal of Clinical Epidemiology, 57, 131-141.	Drug utilization review, the Charlson comorbidity index and the Johns Hopkins Adjusted Care Group Case-Mix profile (ADGs).	Asthma claims were highly specific, Sp= 96.76 (95%CI 96.5, 97.0).

Asthma diagnosis in ambulatory setting	Fowles, J. B., Fowler, E. J., & Craft, C. (1998). Validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. Journal of Ambulatory Care Management, 21(1), 24-34.	Multispecialty group practice in Minneapolis, Minnesota	Sensitivity and specificity was 0.82 and 0.99, respectively. Sensitivity of .82 using claims was higher than sensitivity using self-report at .64
Asthma diagnosis in clinic/outpatient setting	Wilchesky, M., Tamblyna, R. M., & Huang, A. (2004). Validation of diagnostic codes within medical services claims. Journal of Clinical Epidemiology, 57, 131-141.	Drug utilization review, the Charlson comorbidity index and the Johns Hopkins Adjusted Care Group Case-Mix profile (ADGs).	Asthma claims were highly specific, Sp= 96.76 (95%CI 96.5, 97.0).
Bronchitis diagnosis in ambulatory setting	Improving Healthcare for the Common Good (IPRO). Avoidance of Antibiotic Treatment in Adults with Acute Bronchitis. May 2011. http://www.health.ny.gov/health_care/managed_care/reports/docs/adults_antibiotic.pdf	New York Medicaid managed care members	An IPRO analysis of ambulatory claims data in NY State Medicaid found that of 651 individuals with an administrative claim for bronchitis, 629 (96.6%) were confirmed by chart review.
Fill of short acting beta agonist Fill of asthma controller medication	Samnaliev, M., Baxter, J. D., & Clark, R. E. (2009). Comparative evaluation of two asthma care quality measures among Medicaid	Using complete claims and pharmaceutical data for 19,076 patients with persistent asthma (based on Health Effectiveness and Data Information Set criteria) in five Medicaid populations	Sensitivity and specificity were combined into one statistic, the area under the ROC curve. For controller medications, the area under ROC curve is 0.705, which represents good agreement.

• anti- asthmatic combination • antibody inhibitor • inhaled steroid combinations • inhaled corticosteroids (alone or in combination) • leukotriene modifiers • methylxanthines (alone or in combination) • mast cell stabilizers	beneficiaries. Chest, 135(5), 1193-1196. Mudd KE, Bollinger ME, Hus VD, et al. Concordance of Mediaciad and pharmacy record data in inner-city children with asthma. Contemporary Clinical Trials 29(2008) 13-20 Grymonpre R, Xheang M, Fraser M, et al. cvalidity of Precritpion Claims Database to Estimate Medication Adherence in Older Persons e.g. Samnaliev M, Baxter JD, and Clark RE. Comparative Evaluation of Two Asthma Care Quality Measure Among Medicaid Beneficiaries. Berger WE, Legorreta AP, Blaiss MS, et al. The Utility of	(Colorado, Georgia, Indiana, New Jersey, Washington) using ICD-9-CM code 493.x to measure filling prescriptions of asthma control medication. Comparison of pharmacy records and Medicaid clams Manitoba prescription claims and pill count for medication adherence A number of studies found that asthma drug data using the similar HEDIS data elements that we propose were valid for predicting things like emergency department use in asthma patients. As indicated in this article: “HEDIS has become an important industry standard…adopted by	For inner city children on Medicaid, Medicaid claims was sensitive compared to pharmacy records, identifying 91.3% of pharmacy claims for ICS, 94.7% for SABA and 90.4% for leukotriene modifiers (Table 2) Using a much stronger standard of actual compliance, this study found for multiple condition for two conditions in adults that there was strong concordance (79% and 88% respectively) between pill counts and administrative claims data. Not specific for asthma meds Controller medication use was associated with fewer ED visits across 5 states, with OR ranging from 0.30 to 0.47, all significant, overall 0.34 (0.32-0.36). Used actual HEDIS pharmacy code set as do we. Low Controller use had an adjusted odds ratio of 1.72 (1.42-2.08) of ED visit or

	the HEDIS Asthma Measure to predict asthma related outcomes. Annals of Allergy, Asthma, and Immunology. 93:538-545. 2004.	regulators, consumers, and public purchasers of health care" Commercial claims	hospitalization. Those with moderate and higher adherence had graded reductions in undesirable outcomes in the predicted fashion (OR, .84 and 0.72 respectively)
Exclusions			
Diagnosis of COPD	Rawson NS, Malcolm E., validity of the recording of ischaemic heart disease and chronic obstructive pulmonary disease in the Saskatchewan health care datafiles. State Med. 1995. Dec 30: 14 (24):2627-43.	Administrative health care datafiles of the Canadian province of Saskatchewan	Comparisons between hospital data and medical charts for chronic airways obstruction patients showed excellent diagnostic agreement at 94%. In other words, the charted discharge diagnosis from the patient's medical record showed exact agreement for 94.2% of these patients.
	Ginde AA, Tsai CL, Blanc PG, Camargo CA Jr. Positive predictive value of ICD-9-CM codes to detect acute exacerbation of COPD in the emergency department. Jt Comm J Qual Patient Saf.2008;34(11):678–680.	Two academic emergency departments.	The overall positive predictive value for the presence of any of the specified codes, including COPD, was 97%. The positive predictive value for a code of 496 alone was 60% (95% CI 32-84%).
	Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying individuals with physician diagnosed	Claims in Ontario, Canada	The combination of one or more outpatient ICD-9 codes (491.xx, 492.xx, 496.xx) and ICD-10 inpatient ICD-10 codes (J41, J43,

	COPD in health administrative databases. Copd. 2009;6(5):388–394. doi: 10.1080/15412550903140865.		J44) had a sensitivity of 85% and specificity of 78.4% among 113 patients with COPD and 329 patients without COPD.
Diagnosis of COPD Diagnosis of cystic fibrosis Diagnosis of emphysema (Exclusions identified anywhere are excluded. The measure is written to over exclude if need be, but our data suggest that exclusions are uncommon.)	Quan, H., Li, B., Saunders, L. D., Parsons, G. A., Nilsson, C. I., Alibhai, A., et al. (2008). Assessing validity of icd-9-cm and icd-10 administrative data in recording clinical conditions in a unique dually coded database. HSR: Health Services Research, 43(4), 1424.	Four teaching hospitals in Alberta, Canada	Claims had a PPV of 91.9, and a negative predictive value of 92.6, with *k* of 0.65 (substantial agreement[i]) compared to chart review for chronic pulmonary disease. ICD 10 performed similarly in this study
	NCQA: http://www.qualityforum.org/QPS/QPSTool.aspx?m=367&e=1	The presence of diagnostic exclusions was extensively tested on the entire field test population (>82,000 members) to determine the effect on eligible population and the measure results experienced as a result of the application of clinical exclusions.	This measure was deemed valid by the expert panel and approved by NCQA's Committee on Performance Measurement (CPM) for continued inclusion in HEDIS[ii]

Data element	**Reference** (e.g., Quam, et al., 1993)	**Data source** (e.g., Medicare FFS outpatient data)	**Statistical results** (e.g., kappa, sensitivity, specificity, etc.)
Race/ Ethnicity	Kressin, NR, Chang, BH, Hendricks, A, Kazis, LE. Agreement Between Administrative Data and Patients' Self-Reports of Race/Ethnicity. American Journal of Public Health. Oct. 2003. 93 (10): 1734-1739.	Federal administrative data	Among patients with known race/ethnicity, there was a 97.9%, 92.0%, and 83.4% agreement between self-report race/ethnicity and administrative data for white, African American, and Hispanic, respectively. (Table 2, p. 1736)
	Blustein, J. The Reliability of Racial Classifications in Hospital Discharge Abstract Data. American Journal Public Health. 1994; 84:1018-1021.	Statewide Planning and Research Cooperative System, a hospital discharge abstract database maintained by the New York State Department of Health.	Percentage of concordance and kappa of reported racial classifications: - Black: 99%; 089 (95% CI: 0.82, 0.96) - White: 95%; 0.72 (95% CI: 0.64, 0.80) (Table 3, page 1020)
	Klinger, EV, Carlini, SV, Gonzalez, I, et al., Accuracy of Race, Ethnicity, and Language Preference in an Electronic Health Record. 2014. J Gen Intern Med. 30(6):719-23.	Thirteen primary care clinics' electronic health records.	When comparing electronic health record to self-report the sensitivity, specificity and ppv for Black, Hispanic and white are as follows (Table 2, page 721): - Black: Se: 70.9, Sp: 98.8, PPV: 95.5 - Hispanic: Se: 83.8, Sp: 99.8; PPV: 98.9 - White: Se: 93.8; Sp: 97.0; PPV: 98.3
	Escarce, JJ and McGuire, TG., Methods for Using Medicare Data to Compare Procedure Rates among Asians, Blacks, Hispanics, Native Americans, and Whites. Health Services Research. Oct. 2003. 38(5): 1303-1318.	Physician claims data	When comparing enrollment database and survey, probability for White, Black, and Hispanic are 0.954, 0.943, 0.977, respectively. (Table 2, page 1309)

We develop our measure using scientifically sound principles. We first discuss research involving the soundness of our data sources, which include both administrative data to identify cases (and a fraction of numerator qualifications) and chart review (medical record audit) to confirm some denominator inclusions and to identify most numerator inclusion. This is a generally accepted and standard approach with acceptable reliability.

We use administrative data to identify the age of the child, various stratification variables and the presence of asthma, as well as the presence of an asthma ED visit or hospitalization. These are routinely used to support billing by CMS, Medicaid, and private insurers and are routinely used in quality measurement. Administrative data are not typically sufficient for detailed clinical assessment.[1-5] HEDIS developed a hybrid approach, using administrative data and chart review that this measure borrows heavily from. [6, 7]

There is moderate agreement (kappa = 0.45 – 0.50) when comparing administrative data regarding the presence of constructs such as recent asthma attacks, use of asthma medications, attack or medication, attack and medication, using 1 year of administrative claims data. The agreement improves from 0.55 to 0.60 when using two years of data. (8). We expect that these kappas would be significantly higher were the analyses restricted to children with disease that met our construct criteria for identifiable asthma.

The explicit criteria that we use were developed using a slightly modified version of the RAND/UCLA Appropriateness Method that maintained the key aspects of that approach, including a detailed literature review, a multidisciplinary and geographically diverse expert panel comprised of both clinicians and researchers, and the two Round modified Delphi Process. The general reliability of this approach is well established. [9, 10] It has been applied successfully to pediatric services previously. [11-13] We have used as criteria for this measure those specifications whose median rating is 8 or 9, the two highest ratings.

In our testing of the criteria during chart audit used a simple paper data collection instrument that was largely a checklist of yes/no for the various items. After a brief training by the physician who organized the testing three non-clinical research assistants (one MPH, 2 Bachelors) conducted chart audits. Kappa is presented in the next section (**2b2.3)**.

1. Dresser, M.V., et al., *Clinical quality measurement. Comparing chart review and automated methodologies.* Med Care, 1997. **35**(6): p. 539-52.
2. Newton, K.M., et al., *The use of automated data to identify complications and comorbidities of diabetes: a validation study.* J Clin Epidemiol, 1999. **52**(3): p. 199-207.
3. Thompson, B.L., et al., *Measuring clinical performance: comparison and validity of telephone survey and administrative data.* Health Serv Res, 2001. **36**(4): p. 813-25.
4. Angier, H., et al., *Variation in outcomes of quality measurement by data source.*
a. Pediatrics, 2014. **133**(6): p. e1676-82.
5. Weiskopf, N.G. and C. Weng, *Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research.* Journal of the American Medical Informatics Association, 2013. **20**(1): p. 144-151.
6. Pawlson, L.G., S.H. Scholle, and A. Powers, *Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures.* Am J Manag Care, 2007. **13**(10): p. 553-8.
7. NCQA. *National Committee for Quality Assurance*. [cited 2014 7/30/14]; Available from: http://www.ncqa.org/
8. Huzel, L, et al. Diagnosing Asthma: The fit between survey and administrative database. Can Respir J. 2002 Nov-Dec;9(6):407-12.
9. Fitch, K., et al., *The RAND/UCLA Appropriateness Method User's Manual*. 2001 RAND.
10. Kosecoff, J., et al., *The appropriateness of using a medical procedure. Is information in the medical record valid?* Med Care, 1987. **25**(3): p. 196-201.
11. Kleinman, L.C., et al., *The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States.* Jama, 1994. **271**(16): p. 1250-5.
12. Kleinman, L.C., E.A. Boyd, and J.C. Heritage, *Adherence to prescribed explicit criteria during utilization review. An analysis of communications between attending and reviewing physicians.* Jama, 1997. **278**(6): p. 497-501.
13. Keyhani, S., et al., *Overuse of tympanostomy tubes in New York metropolitan area: evidence from five hospital cohort.* Bmj, 2008. **337**: p. a1607.

Table 1. 360 Degree Pediatric Quality Measure Development: Overview

Stage	Phase	Innovation	Product(s)
1. Clinical Criteria Development	a. Input Development	1. Focus groups of caregivers of children with asthma who have used the ED 2. Interviews with front line clinicians: primary care, asthma docs, and ED docs	1. Literature review 2. Summary of consumer perspectives, values and understanding relevant to clinical issue of interest 3. Summary of findings form clinician interviews
	b. RAND/UCLA 2 Round Modified Delphi Process	1. Inclusion of consumer perspectives as a key input; 2. Use of this method to identify appropriateness criteria in national performance measure development;	1. Explicit criteria that rank a comprehensive and mutually exclusive set of clinically detailed scenarios;
2. Boundary Guideline Development	Criteria Enhancement	1. Iterative process to enhance reliability and internal consistency of the explicit criteria set with a goal of outlining three boundary spaces	1. Internally consistent set of explicit criteria that are stable in their representation of the expert panel perspective. "Enhanced criteria"
	Guideline Articulation	1. Stakeholder (including experts, users, clinicians, consumers and others) informed review of the enhanced criteria. 2. Definition of zones of potential overuse, potential underuse, and professional interaction and decision-making based upon the explicit criteria 3. Stakeholder valuations of potential deviations from guideline 4. Boundary Guideline	1. Boundary Guideline 2. Prioritization list
3. Creation of Measure	Specification	1. Translation of guideline into specification of necessary data 2. Iterative process to define optimally efficient sources of data to allow for measurement and stratification	1. Initial specification of measure
	Review	1. Constructive peer review of specifications by stakeholders in Steering Committee and SAB	1. Final specifications of measure including variables for stratification as needed
	Fielding and testing of measure	1. Measure testing	1. Functional experience and practical understanding of measure, its scoring, variability, and interpretation

This measure was developed and assessed using a pre-specified process and consistent with CAPQuaM's peer reviewed 360 degree method outlined in the table above.
Explicit criteria were developed using a variation of the two-round modified Delphi process RAND/UCLA Appropriateness Method with a multidisciplinary and geographically diverse expert panel comprised of both clinicians and researchers. Identifiable asthma was based on panel findings and appropriateness criteria included for this measure were those that were both available in the chart and highly rated.

Development included a series of alpha tests to refine specifications by conducting iterative analyses in New York State Medicaid data. Conclusions from alpha tests include:

1) The reporting period and the assessment period could not overlap completely, leading to use of 2 years of data as shown in the specifications' diagram. The optimal approach was to divide the reporting year into 12 reporting months. ED events in that month are eligible for the numerator if persistent asthma criteria have been satisfied (combining the look-back year and all prior months in the reporting year) and the child has been continuously

enrolled for the two months immediately prior to the reporting month. The optimal unit for the denominator is in child-months;

2) Using both revenue codes and CPT codes increased our sensitivity meaningfully, a choice validated by consultation with coding and billing experts;
3) NY State Medicaid data and national survey data (HCUP) converged to demonstrate the importance of including hospitalizations as numerator events even when the underlying construct is ED visits. This is consistent with policies of many payers to request providers not to submit both ED and hospital claims for the same day. Error would be far less by considering both ED visits and hospitalizations as numerator events, than by not including hospitalizations.
4) The expert panel only wanted numerator events for which the children were already known to the accountable entity as having asthma and established definitions for such "identifiable asthma". Identifiable asthma was intended to be more restrictive than the 15-16% identified by our analysis of the 2011 NSCH as having ever been told they had asthma and much less restrictive than the HEDIS definition of persistent asthma. Alpha testing in NY State Medicaid demonstrated the expected results:
    a. Holding steady the continuous enrollment criterion at 12 months, HEDIS criteria identified a rate of persistent asthma of 3.1%, the CAPQuaM criteria identifying identifiable asthma at a rate of 8.6%. This ratio is 2.8 (our predicted and target result was between 2-3 based the literature achieve and our intended construct).
    b. Relaxing the continuous enrollment period to 3 months was suggested by members of our stakeholder steering committee. Doing so increased the eligible number by several tens of thousands while still restricting the measure to those who had received sufficient care for asthma to be identified, and requiring continuous enrollment for attribution to the extent felt important by our multi-stakeholder group.

The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children. [1] Use of the medical record as a valid source of information to judge appropriateness is well accepted. [2] Chart audits are used frequently to generate research in Emergency Medicine. [3, 4]

Key panel ratings are shown. Constructs rated 7 or higher are endorsed, 8 or higher strongly endorsed, and 2 or lower strongly rejected.

Scenario	**MED**
Wheezing on presentation to the ED establishes that the ED was an appropriate level of care for that child.	5
Retractions or labored breathing during the ED visit establishes that the ED was an appropriate level of care for that child.	**9**
Decreased breath sounds establish that the ED was an appropriate level of care.	6
Markedly decreased breath sounds establish that the ED was an appropriate level of care.	**7**
Obtaining an ABG in the ED establishes the ED as an appropriate level of care for that child.	**9**
Oxygen saturation less than 90% establishes that the ED was an appropriate level of care for that child.	**9**
Hospitalization following the ED visit establishes that the ED was an appropriate level of care for that child.	**9**
An ED visit less than 72 hours following a previous ED visit in a child with asthma establishes that the ED was an appropriate level of care for that child.	4
Prescription of an oral steroid burst establishes that the ED was an appropriate level of care for that child.	4
An ED visit less than one week following a hospital discharge in a child with asthma establishes that the ED was an appropriate level of care for that child.	4
An ED visit less than 72 hours following a hospital discharge in a child with asthma establishes that the ED was an appropriate level of care for that child.	3
A specialty consultation in the ED establishes that the ED was an appropriate level of care for that child.	**8**
Homelessness establishes that the ED was an appropriate level of care for that child.	3
Parent report that the PCP is generally unavailable for urgent asthma care establishes that the ED was an appropriate level of care for that child.	5
Parent report of inability to reach the PCP during the current event establishes that the ED was an appropriate level of care for that child.	6

Parent report that they were referred into the ED by phone contact a clinician establishes that the ED was an appropriate level of care for that child.	**8**
Parent report that they were referred to the ED after being seen by a clinician establishes that the ED was an appropriate level of care for that child.	**9**
Parent report that the child did not respond to a dose of a rescue medication establishes that the ED was an appropriate level of care for that child.	6
Parent report that they are unable to afford needed asthma medications establishes that the ED was an appropriate level of care for that child.	3
Parent report that they are unable to obtain needed care because of financial barriers establishes that the ED was an appropriate level of care for that child.	3

The proportion of visits found to be appropriate varied by age and there are biological reasons that make plausible such differences not only being related to health services. Therefore we have specified this measure to be reported as stratified by age. Our data showed that within the 2 – 5 year age group, 54.3% were appropriate, within the 6 – 11 year age group, 44.3% were appropriate and within the 12 – 18 year age group, 48.3% were appropriate, p =.019. The breakdown is as follows:

- For children 2-5: 181 of 335 audits (54.3%) were deemed appropriate.
- For children 6-11: 209 of 477 audits (43.8%) were deemed appropriate.
- Adolescents aged 12-18: 165 of 341 audits (48.4%) were deemed appropriate based upon information in the chart audit.

Criteria for appropriateness that were met were recorded and did vary by age.

1. Brook, R.H., et al., *A method for the detailed assessment of the appropriateness of medical technologies.* International journal of technology assessment in health care, 1986. **2**(01): p. 53-63.
2. Kosecoff, J., et al., *The appropriateness of using a medical procedure: is information in the medical record valid?* Med Care, 1987: p. 196-201.
3. Gilbert, E.H., et al., *Chart reviews in emergency medicine research: where are the methods?* Ann Emerg Med, 1996. **27**(3): p. 305-308.
4. Worster, A., et al., *Reassessing the methods of medical record review studies in emergency medicine research.* Ann Emerg Med, 2005. **45**(4): p. 448-51.

The development team's goal was to develop an ICD10 code set that was fully consistent with the intent of the original measure. Our process began by performing general equivalency mapping using the forward mapping from www.icd9data.com. We then did a de novo review of the CMS ICD 10 CM set to seek to identify codes that might be appropriate for asthma. We reviewed potential codes identified by both sources and developed a new list of codes appropriate for inclusion criteria and a new list of codes appropriate for exclusion criteria. Drs. Kleinman and Sharma reviewed the lists independently and then achieved consensus in a conference call review and discussion. Key team members for this work were Suzanne Lo, MPH who staffed and coordinated this work, Sandeep Sharma, MD, Dr.PH and Lawrence Kleinman, MD, MPH. Dr. Sharma was a lead developer for one of CAPQuaM's 2 asthma measures and Dr. Kleinman is both CAPQuaM PI and was a lead developer for both measures. The guidance for the intended constructs for both ICD9 and ICD10 coding were the findings from a RAND style modified Delphi panel that incorporated 9 national experts over the course of the measure development process.

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

We assessed the reliability of data abstraction. Three reviewers each reviewed 10 charts early in training and after further practice. This results in 180 comparisons with the trainer (6 clinical constructs * 3 * 10 = 180). Another 30 comparisons may be made based upon the global scoring chart review as appropriate if any criterion was met. The 6 constructs were findings from the chart review : retractions; accessory muscles being used; markedly reduced breath sounds; ED visit resulted in hospitalization, oxygen saturation was documented in the ED below 90%; PCP referred the patient into the ED. In our pretesting we found that accessory muscle use did not enhance the sensitivity of the appropriateness construct, so we integrated the two with grunting into a category of labored breathing.

	Agreement	
Construct	**Initial Kappa**	**Final Kappa**

7. Retractions	0.67	0.87
8. Accessory Muscle Use	0.44	0.89
9. Markedly diminished BS	0.71	0.78
10. Hospitalized from ED	1.0	1.0
11. O2 sat < 90%	0.79	NA*
12. Referred by PCC	1.0	NA*
All six combined	0.76	0.68
Overall: Appropriateness	0.77	0.87

* NA is because there was no variability in the charts reviewed. There was no disagreement in any of the assessments.

The key assessment (since it is the bottom line) is agreement regarding appropriateness, which is the highlighted row near the bottom of the table. We found that after training and with practice kappa moved from an already strong 0.77 to an excellent 0.87, confirming excellent reliability at the level of the numerator. We confirm that non-medically trained research assistants can be trained to do sufficient quality data abstractions to assess appropriateness of ED visits using these specifications.

IDENTIFYING A POPULATION WITH ASTHMA:

For the foundational NCQA work, NCQA's field test retested a number of previously validated criteria for identifying an eligible population with persistent asthma using administrative claims data. Using the dataset provided, NCQA examined several different scenarios to determine the effects of different specification criteria on this particular population. This information was combined with multiple years of HEDIS data collection of this measure to examine the reliability of collecting this measure through administrative claims.

Score level reliability of the HEDIS 2011 submissions (2010 data) was assessed using the beta-binomial model. Beta-binomial is a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® health plan measures. The beta-binomial model assumes the plan score is a binomial random variable conditional on the plan´s true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates. The beta distribution can be symmetric, skewed or even U-shaped.

Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

We cite these not as specific evidence of score level performance of the submitted measure, but as evidence that the HEDIS measures that rely on the same administrative data elements for their denominator have the capacity to distinguish signal to noise at a very high level. If the population assessment were inadequate, then these other measures which use the same data elements to establish their denominators could not achieve such high reliability scores. This is because failure to distinguish signal from noise at the level of the HEDIS denominators would lead to non-differential misclassification error which is a major bias towards the null, in other towards noise and away from signal. Hence these provide strong indirect evidence of the validity of our approach to capturing the measure's denominator.

While there is moderate agreement (kappa = 0.45 - 0.50) when comparing administrative data regarding the presence of constructs such as recent asthma attacks, use of asthma medications, attack or medication, attack and medication, using

1 year of administrative claims data to parent report, the agreement improves from 0.55 to 0.60 when using two years of data. (Huzel, L. et al. Diagnosing Asthma: The fit between survey and administrative database. Canada Resp. Journal 2002.) We expect that these kappas would be significantly higher were the analyses restricted to children with disease that met our construct criteria for identifiable asthma.

Further, we identify asthma visits and medications using the same data that an insurance company or Medicaid would use for payment, including ICD9 codes, CPT codes, and revenue codes. We have had conversations with expert coders and New York State Department of Health Office of Health Insurance Programs to confirm our choices.

The literature also supports our work. ICD-09 and ICD-10 codes for asthma on patients' medical charts typically match claims data. ICD-9-CM administrative data have been validated using various methodologies for various purposes (5-17). Studies have shown high sensitivity and specificity for diagnoses obtained from administrative data among children with high-risk conditions including asthma, (18), and high predictive value among adolescents and adults with asthma. (19) (20) HEDIS criteria using administrative data support peer reviewed research, for example in patients with persistent asthma based on HEDIS criteria in five Medicaid programs (Colorado, Georgia, Indiana, New Jersey, Washington) using ICD-9-CM code 493.x (21). Fowles and colleagues report sensitivity and specificity of claims compared with ambulatory medical records to identify asthma was 0.82 and 0.99, respectively. (22) Wilchesky compared chart abstraction to diagnoses obtained from administrative database: asthma claims were highly specific, Sp= 96.76 (95%CI 96.5, 97.0). (23) Bronstein et al found that 88.3% of diagnoses asthma on claims agreed with medical record, with a negative predictive value of 0.85 and a positive predictive value of 0.88.They conclude that claims are generally an accurate indicator of the content of a patient encounter. (24) Steinwachs et al. compared billed claims to medical records based on date of visit and diagnosis, they found for asthma there was 90.9 percent of billed visits in record on same date and 82.8 percent of billed visits with same diagnosis in record on same date. (25) Quan et al documented the validity of ICD-9-CM and ICD-10 coding systems in coding clinical information and found that ICD-10 data was generally comparable with that of ICD-9-CM data in recording clinical information. (26)

From a public health perspective, asthma surveillance systems in several states, including Maine, North Carolina, Connecticut and Michigan, have shown the feasibility of using administrative data to identify children having asthma, based on primary and secondary diagnosis codes reported on inpatient and outpatient claims. (27-30) Researchers also classified children with evidence of persistent asthma using HEDIS criteria, (31). Another study showed the usefulness of ICD9 493.x to identify asthma for a quality measure using Maryland data. Like our measure, those researchers excluded children with a diagnosis of cystic fibrosis (ICD9 277). (32) regarding our capacity to identify exclusions, Quan et al found that claims had a PPV of 91.9, and a negative predictive value of 92.6, with *k* of 0.65 (substantial agreement[1]) compared to chart review. FICD 10 performed similarly in this study.

REFERENCES

(1) National Committee for Quality Assurance. (1 994). Health care quality improvement studies in managed care set1ing.s: A guide for state medicaid agencies. Washington, DC: Author.

(2) Leatherman, S., Peterson, E., Heinen, L., & Quam, L. (1991). Quality screening and management using claims data in a managed care setting. Quality Review Bulletin, 17, 349-359.

(3) Parente, S.T., Weiner,J.P., Garnick, D.W., Richards, T.M., Fowles,J., Lawthen, A.G., Chandler, P., & Palmer, R.H. (1995). Developing a quality improvement database using health insurance data: A guided tour with applications to Medicare's National Claims History File. American Journal of Medical Quality, 10(4), 162-176.

(4) Weiner, J.P., Powe,. R., Steinwachs, D.M., & Dent, G. (1990). Applying insurance claims data to assess quality of care: A compilation of potential indicators. Quality Review Bulletin, 16, 424-438.

(5) Best, W.R., S.F. Khuri, M. Phelan, K. Hur, W.G. Henderson, J.G. Demakis, and J. Daley. 2002. ''Identifying Patient Preoperative Risk Factors and Postoperative Adverse Events in Administrative Databases: Results from the Department of Veterans Affairs National Surgical Quality Improvement Program.'' Journal of American College of Surgeons 194: 257–66.

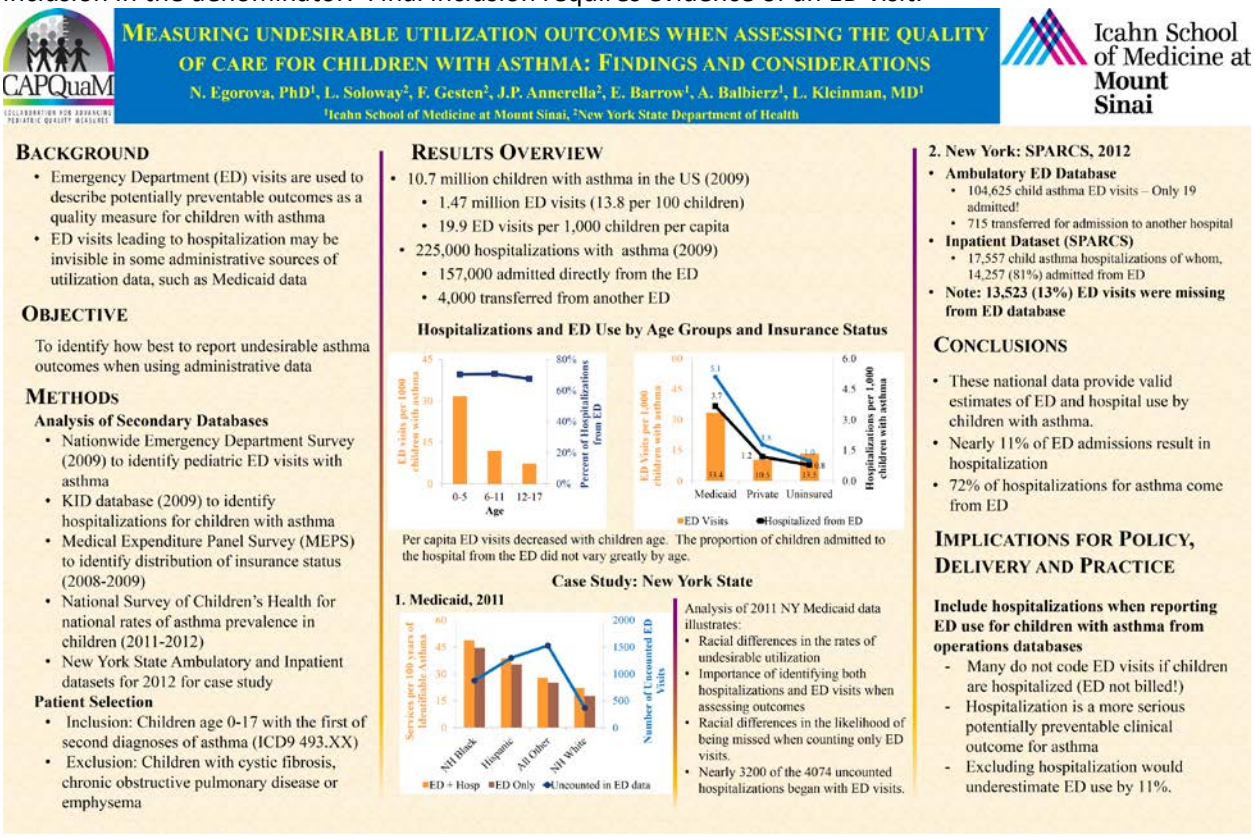
(6) Geraci, J.M., C.M. Ashton, D.H. Kuykendall, M.L. Johnson, and L. Wu. 1997. ''International Classification of Diseases, 9th Revision, Clinical Modification Codes in Discharge Abstracts are Poor Measures of Complication Occurrence in Medical Inpatients.'' Medical Care 35: 589–602.

(7) Hsia, D.C., C.A. Ahern, B.P. Ritchie, L.M. Moscoe, and W.M. Krushat. 1992. ''Medicare Reimbursement Accuracy under the Prospective Payment System, 1985 to 1988.'' Journal of the American Medical Association 268: 896–9.

(8) Iezzoni, L.I., S. Burnside, L. Sickles, M.A.Moskowitz, E. Sawitz, and P.A. Levine. 1988. ''Coding of Acute Myocardial Infarction. Clinical and Policy Implications.'' Annals of Internal Medicine 109: 745–51.
(9) Jollis, J.G., M. Ancukiewicz, E.R. DeLong, D.B. Pryor, L.H. Muhlbaier, and D.B. Mark. 1993. ''Discordance of Databases Designed for Claims Payment versus Clinical Information Systems. Implications for Outcomes Research.'' Annals of Internal Medicine 119: 844–50.
(10) Lee, D.S., L. Donovan, P.C. Austin, Y. Gong, P.P. Liu, J.L. Rouleau, and J.V. Tu. 2005. ''Comparison of Coding of Heart Failure and Comorbidities in Administrative and Clinical Data for Use in Outcomes Research.'' Medical Care 43: 182–8.
(11) Muhajarine, N., C. Mustard, L.L. Roos, T.K. Young, and D.E. Gelskey. 1997. ''Comparison of Survey and Physician Claims Data for Detecting Hypertension.'' Journal of Clinical Epidemiology 50: 711–8.
(12) Quan, H., G. Parson, and W. Ghali. 2002. ''Validity of Information on Comorbidity Derived from ICD-9-CM Administrative Data.'' Medical Care 40: 675–85.
(13) Romano, P.S., B.K. Chan, M.E. Schembri, and J.A. Rainwater. 2002. ''Can Administrative Data Be Used to Compare Postoperative Complication Rates across Hospitals?'' Medical Care 40: 856–67.
(14) Romano, P.S., and D.H. Mark. 1994. ''Bias in the Coding of Hospital Discharge Data and Its Implications for Quality Assessment.'' Medical Care 32: 81–90.
(15) Weingart, S.N., L.I. Iezzoni, R.B. Davis, R.H. Palmer, M. Cahalane, M.B. Hamel, K. Mukamal, R.S. Phillips, D.T. Davies Jr., and N.J. Banks. 2000. ''Use of Administrative Data to Find Substandard Care: Validation of the Complications Screening Program.'' Medical Care 38: 796–806.
(16) Yasmeen, S., P.S. Romano, M.E. Schembri, J.M. Keyzer, and W.M. Gilbert. 2006. ''Accuracy of Obstetric Diagnoses and Procedures in Hospital Discharge Data.'' American Journal of Obstetrics and Gynecology 194: 992–1001.
(17) Quam, L., Ellis, L. B. M., Venus, P., Clouse, J., Taylor, C. G., & Leatherman, S. (1993). using claims data for epidemiologic research. the concordance of claims-based criteria with the medical recordand patient survey for identifying a hypertensive population. Medical Care, 31(6), 498-507.
(18) Daley MF, Barrow J, Pearson K, Crane LA, Gao D, Stevenson JM, et al. Identification and recall of children with chronic medical conditions for influenza vaccination. Pediatrics 2004;113(1 Pt 1):e26-33.
(19) Vollmer WM, O'Connor EA, Heumann M, Frazier EA, Breen V, Villnave J, Buist AS. Searching multiple clinical information systems for longer time periods found more prevalent cases of asthma. J Clin Epidemiol 2004;57:392-7.
(20) Twiggs, J. E., Fifield, J., Apter, A. J., Jackson, E. A., & Cushman, R. A. (2002). Stratifying medical and pharmaceutical administrative claims as a method to identify pediatric asthma patients in a medicaid managed care organization. Journal of Clinical Epidemiology, 55(9), 938-944.
(21) Samnaliev, M., Baxter, J. D., & Clark, R. E. (2009). Comparative evaluation of two asthma care quality measures among medicaid beneficiaries. Chest, 135(5), 1193-1196.
(22) Fowles, J. B., Fowler, E. J., & Craft, C. (1998). validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. Journal of Ambulatory Care Management, 21(1), 24-34.
(23) Wilchesky, M., Tamblyna, R. M., & Huang, A. (2004). Validation of diagnostic codes within medical services claims. Journal of Clinical Epidemiology, 57, 131-141.
(24) Bronstein, J. M., Santer, L., & Johnson, V. (2000). The use of medicaid claims as a supplementary soure of information on quality of asthma care. Journal for Healthcare Quality, 22(6), 13-18.
(25) Steinwachs, D. M., Stuart, M. E., Scholle, S., Starfield, B., Fox, M. H., & Weiner, J. P. (1998). A comparison of ambulatory medicaid claims to medical records: A reliability assessment. American Journal of Medical Quality, 13(2), 63-69.
(26) Quan, H., Li, B., Saunders, L. D., Parsons, G. A., Nilsson, C. I., Alibhai, A., et al. (2008). Assessing validity of icd-9-cm and icd-10 administrative data in recording clinical conditions in a unique duallycoded database. HSR: Health Services Research, 43(4), 1424.
(27) Deprez RD, Asdigian NL, Oliver LC, Anderson N, Caldwell E, Baggott LA. Development of a prototype system for statewide asthma surveillance. Am J Public Health 2002;92:1946-51.
(28) Buescher PA, Jones-Vessey K. Using Medicaid data to estimate state- and county-level prevalence of asthma among low-income children. Matern Child Health J 1999;3:211-6.
(29) Children's Health Council. Asthma affects nearly one in ten children enrolled in HUSKY Part A in 1999–2000. Hartford (CT): Children's Health Council; 2001.
(30) Dombkowski, K. J., Wasilevich, E. A., & Lyon-Callo, S. K. (2005). Pediatric asthma surveillance: Using medicaid claims. Public Health Reports, 120, 515.
(31) National Committee for Quality Assurance. Use of appropriate medications for people with asthma. HEDIS 2003, volume 2: technical specifications. Washington: NCQA; 2003.
(32) Stuart, B., Singhal, P. K., Magder, L. S., & Zuckerman, I. H. (2003). How robust are health plan quality indicators to data loss? A monte carlo simulation study of pediatric asthma treatment. HSR: Health Services Research, 38(6), 1547.
(33) Schneeweiss, S., & Avorn, J. (2005). A review of uses of health care utilization databases for epidemiologic research on therapeutics. Journal of Clinical Epidemiology, 58(4), 323-337.

Our own research looking at NY State Medicaid and national all payer data (see poster presented at peer-reviewed AcademyHealth national meeting) is consistent with expert and other recommendations that to identify all ED visits, one also needs to include hospitalizations for asthma as potential indicators of an otherwise unrecognized ED visit, which we have done and incorporated into the specifications.

This is the poster presenting our original research regarding the inclusion of hospitalizations when considering potential inclusion in the denominator. Final inclusion requires evidence of an ED visit.



MEASURING UNDESIRABLE UTILIZATION OUTCOMES WHEN ASSESSING THE QUALITY OF CARE FOR CHILDREN WITH ASTHMA: FINDINGS AND CONSIDERATIONS

N. Egorova, PhD[1], L. Soloway[2], F. Gesten[2], J.P. Annerella[2], E. Barrow[1], A. Balbierz[1], L. Kleinman, MD[1]

[1]Icahn School of Medicine at Mount Sinai, [2]New York State Department of Health

Icahn School of Medicine at Mount Sinai

BACKGROUND

- Emergency Department (ED) visits are used to describe potentially preventable outcomes as a quality measure for children with asthma
- ED visits leading to hospitalization may be invisible in some administrative sources of utilization data, such as Medicaid data

OBJECTIVE

To identify how best to report undesirable asthma outcomes when using administrative data

METHODS

Analysis of Secondary Databases

- Nationwide Emergency Department Survey (2009) to identify pediatric ED visits with asthma
- KID database (2009) to identify hospitalizations for children with asthma
- Medical Expenditure Panel Survey (MEPS) to identify distribution of insurance status (2008-2009)
- National Survey of Children's Health for national rates of asthma prevalence in children (2011-2012)
- New York State Ambulatory and Inpatient datasets for 2012 for case study

Patient Selection

- Inclusion: Children age 0-17 with the first of second diagnoses of asthma (ICD9 493.XX)
- Exclusion: Children with cystic fibrosis, chronic obstructive pulmonary disease or emphysema

RESULTS OVERVIEW

- 10.7 million children with asthma in the US (2009)
  - 1.47 million ED visits (13.8 per 100 children)
  - 19.9 ED visits per 1,000 children per capita
- 225,000 hospitalizations with asthma (2009)
  - 157,000 admitted directly from the ED
  - 4,000 transferred from another ED

Hospitalizations and ED Use by Age Groups and Insurance Status

Per capita ED visits decreased with children age. The proportion of children admitted to the hospital from the ED did not vary greatly by age.

Case Study: New York State

1. Medicaid, 2011

Analysis of 2011 NY Medicaid data illustrates:

- Racial differences in the rates of undesirable utilization
- Importance of identifying both hospitalizations and ED visits when assessing outcomes
- Racial differences in the likelihood of being missed when counting only ED visits.
- Nearly 3200 of the 4074 uncounted hospitalizations began with ED visits.

2. New York: SPARCS, 2012

- **Ambulatory ED Database**
  - 104,625 child asthma ED visits – Only 19 admitted!
  - 715 transferred for admission to another hospital
- **Inpatient Dataset (SPARCS)**
  - 17,557 child asthma hospitalizations of whom, 14,257 (81%) admitted from ED
- **Note: 13,523 (13%) ED visits were missing from ED database**

CONCLUSIONS

- These national data provide valid estimates of ED and hospital use by children with asthma.
- Nearly 11% of ED admissions result in hospitalization
- 72% of hospitalizations for asthma come from ED

IMPLICATIONS FOR POLICY, DELIVERY AND PRACTICE

Include hospitalizations when reporting ED use for children with asthma from operations databases

- Many do not code ED visits if children are hospitalized (ED not billed!)
- Hospitalization is a more serious potentially preventable clinical outcome for asthma
- Excluding hospitalization would underestimate ED use by 11%.

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

Our interpretation is that administrative data are reliable and valid for identifying asthma, and that year to year test retest reliability seems to indicate similar patterns of performance when identifying ED visits for asthma, reinforcing the reliability of our operational definitions for identifying eligible children. Our specification provide a sensitive and face valid approach to identifying an unbiased sample of children with ED visits(ensuring we don't bias the results towards the inappropriate by missing those with hospitalization).

Most databases contain consistent elements, are available in a timely manner, provide information about large numbers of individuals, and are relatively inexpensive to obtain and use. Validity of many databases has been established, and their strengths and weaknesses relative to data abstracted from medical records and obtained via survey have been documented (30). Administrative data are supported, if not encouraged by federal agencies, such as NIH, AHRQ, HCFA, and the VA. The Centers for Medicare & Medicaid Services has made clear to the participating AHRQ-CMS CHIPRA Centers of Excellence funded to develop measures in the Pediatric Quality Measures Program that it places a premium on feasibility when assessing those measures that it will most highly recommend to states to complete. The sources of data for the existing measure and other similar measures are typically based upon administrative data as well, providing consensual validation for using administrative data as the primary data source.

Our Kappa results indicated excellent agreement in the reliability of the chart audit. Kappa values over 0.75 are considered excellent, 0.40 to 0.75 as fair to good, and below 0.40 as poor.

2b3. EXCLUSIONS ANALYSIS

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

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

Exclusions were only included if they were endorsed by the expert panel. In studying the denominator we found that a very few percent of potentially eligible children (<=2.5%) were excluded by clinical diagnoses. The use of three months of continuous enrollment was recommended by our multi-stakeholder consortium and avoids the exclusion of more than 20% of otherwise eligible children from the population with identifiable asthma compared to a 12 month requirement.

Denominator Exclusions
Children with concurrent or pre-existing: Chronic Obstructive Pulmonary Disease (COPD) diagnosis (ICD-9 Code: 496), Cystic Fibrosis diagnosis (ICD-9 code 277.0, 277.01. 277.02, 277.03, 277.09), or Emphysema diagnosis (ICD-9 code 492xx).

Children who have not been consecutively enrolled in the reporting plan for at least two months prior to the index reporting month, as well as the index reporting month itself.

There are no numerator exclusions.

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

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

Exclusions are clinical and represent construct validity rather than statistical considerations.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

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

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☒ **Stratification by** 1 **risk categories**

☒ **Other, There are optional categories for stratification of outcomes, such as race/ethnicity that are for descriptive and not risk stratification purposes. The NHLBI guideline clearly articulates a preference for no such stratifications based upon race/ethnicity, insurance, etc.**

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

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

Specifications for this measure requires stratification by age group. Several additional stratifications are optional but may be required by the accountability entity or provided by the reporting entitity. These variables include race/ethnicity, rurality/urbanicity and county level of poverty.

Within age group, we specify a number of stratifications as we have done for all of our CAPQuaM PQMP measure. Absent clear biological evidence that ED visits should be more likely in any of the sub categories we have chosen not to adjust but to report both topline and stratified results.

The NIH NHLBI NAEPP (http://www.nhlbi.nih.gov/health-pro/guidelines/current/asthma-guidelines/full-report) guideline notes that goals of care and definition of successful management are the same regardless of baseline presentation. Hence clinical risk adjustment is not appropriate.

(page 38)

> ***"An important point linking asthma severity, control, and responsiveness is that the goals are identical for all levels of baseline asthma severity.*** *A patient who has severe persistent asthma compared to a patient who has mild persistent asthma, or a patient who is less responsive to therapy may require more intensive intervention to achieve well-controlled asthma; however, the goals are the same: in well-controlled asthma, the manifestations of asthma are minimized by therapeutic intervention."*

High levels of appropriateness suggest that the children in the ED are there because of an immediate clinical need and the ED service is well utilized. Some of these may have been preventable with better quality care prior to the ED visit and some will not. When appropriateness sis high, Asthma ED visit rates represent a strong proxy for asthma clinical outcomes.

Low levels of appropriateness suggest that the cause of many ED visits is not break through asthma or failures of biological asthma management, but insufficient access or quality of care provided that families are seeking care in the ED as preferential to a less acute setting. The good news in such a finding is that clinical asthma outcomes are better than would appear simply by counting the number of ED visits.

The results have independent meaning but from both accountability and improvement perspectives there is synergy in the interpretation of this measure with the CAPQuaM rate of ED visits in asthma measure.

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

The conceptual model is that of CAPQuaM that includes that in pediatrics age is a key predictor and stratification is valuable. We were asked by AHRQ and CMS to include other constructs and we have manifest them as specified, such as race/ethnicity, poverty level in the caregivers county of residence, rurality/urbanicity on the caregiver's county of residence, insurance type and plan type, when variable. We have not added a stratum for children with special health care needs since asthmatics going to the emergency room are highly likely to belong in this category.

2b4.4a. What were the statistical results of the analyses used to select risk factors?
There are statistically significant differences by age group.

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

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

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

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

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

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

2b4.9. Results of Risk Stratification Analysis:

For results of age-stratified analysis, please refer to section 2b4.4a

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

Our medical chart audit found that the measure varies by race/ethnicity.

Appropriateness varied by age (Chi^2=8.2,p=.02), with younger (p=.01) and school aged (p=.01) children each being significantly different; Adolescents experienced a level of appropriateness intermediate to the other two groups and were not significantly different from them when combined (ie comparing Adolescents to All others). We also found racial differences with Hispanics at 44.1% appropriateness, non-Hispanic Blacks at 51.3%, Whites at 56.5% and all others at 72.2%. Chi square with 3 degrees of freedom was 15.4, with p=.0015. The appropriateness of ED visits for Hispanic children was less than for other children (p=.002).

Hispanic children had higher rates of questionable use of the ED (55.9% of visits) when compared to non-Hispanic children (46.8%), p=.002. Black children showed a trend toward more questionable use compared to all other children (53.6% questionable vs 48.7%, p=.10).

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

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

Contingency table analysis with chi square, SAS 9.4 Generalized linear models (Proc GLM) and SAS 9.4 Logistic Regression (Proc Logistic) analyses were performed and were coherent and each illustrated the presence of statistical differences among identifiable subgroups. All of this work was done in a single hospital facility.

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

The proportion of visits found to be appropriate varied by age and there are biological reasons that make plausible such differences not only being related to health services. Therefore we have specified this measure to be reported as stratified by age. Our data showed that:

- For children 2-5: 181 of 335 audits (54.3%) were deemed appropriate.
- For children 6-11: 209 of 477 audits (43.8%) were deemed appropriate.
- Adolescents aged 12-18: 165 of 341 audits (48.4%) were deemed appropriate

based upon information in the chart audit.

Criteria for appropriateness that were met were recorded and did vary by age

The GLM models regressed appropriateness simultaneously on the class variables Age Group, Ethnicity, Gender, and presence or absence of private insurance found that gender (P=.017), Hispanic ethnicity (p=.002), and private insurance (p=.005) were all significantly associated with level of appropriateness, as was age group (p=.009). For this analysis, N=1,188 with a model F value of 6.56 (Pr>F is <0.0001).

To confirm the distinction between what we expected to be strong and weak effects, we substituted day of week for the various demographic variables other than age group. The P value for day of week (as a class variable) was >0.30. The non-zero effect size is consistent with social science literature that suggests that variables such as time of day and day of week are weakly meaningful. Still, the lack of a significant finding in a reasonably good-sized data set demonstrates that spurious significant findings are not likely to be identified as significant.

Differences between major subgroups were statistically significant, including race/ethnicity, age group, and insurance status. We note that this is a stricter test than had the measure been assessed across different entities. These data showed differences by type of insurance, which in this case can serve as proxy for health plan. The F Value for Insurance Status (after controlling for age group) was 3.91 with 4 degrees of freedom, which exceeded the critical value and is associated with a p-value <0.004. This should correlate with excellent capacity to distinguish between health plans.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

In sum, we found true signal in social determinants (consistent with the asthma literature) and did not incorrectly identify weak signal as meaningful. The measure distinguishes signal from noise.

The measures are sensitive enough to detect meaningful differences as observed within a population (as described above). Since the sum of squares across populations is expected to be greater in distinct populations, we expect the measure to perform very well when comparing across populations as well. Since the effective sample size of within population comparisons (such as we have conducted) is diminished by an (unmeasured) intraclass correlation coefficient, we would expect greater power for equal sample size to detect differences between entities than we had in our testing of various subpopulations within a single state. This supports the same conclusion. The signal to noise ratio is very strong for this measure.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

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

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

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

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

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

Chart review data has been shown to be an accurate method for identifying the presence or absence of conditions required to identify the level of appropriateness of a clinical service. Documentation is a part of the clinical responsibility and failure to document is a quality deficit that is not construed as missing data. Since inclusion requires the affirmative presence of data and we are unaware of any evidence to suggest that there would be differential absence of data between appropriate and non-appropriate visits we are not concerned about introducing bias in our findings. Further, we use random sampling of eligible visits as another means to avoid the introduction of bias. (1, 2)

1. Kahn, KL, Kosecoff, J, Chassin, MR, et al. Measuring the clinical appropriateness of a procedure: Can we do it? Medical Care, 1988, 26:415-422.
2. Kosecoff, J., et al., The appropriateness of using a medical procedure. Is information in the medical record valid? Med Care, 1987. **25**(3): p. 196-201.

While assessing the definition for identifiable asthma, our colleagues at NY State Medicaid conducted a series of iterative analysis using NY State Medicaid Managed care data to assess the importance of our data elements and definitions. These analyses helped to confirm the importance of using, for example, both revenue codes and procedure codes to identify ED visits. These analyses also confirmed that the use of pharmaceutical data to identify children with asthma expanded the pool of these so identified and quantified that statewide doing so added around 10,000 to a total

of around 190,000 children with identifiable asthma in the state. We found no evidence that this was a threat to the measure's validity. The key reason for inclusion of pharmacy data is that our expert panel directed us to use it and it is a slightly more sensitive way to identify asthmatic children from the pool of all children with asthma related claims. The expert panel did not want the absence of pharmacy data to preclude inclusion of a reporting entity in the measure or to exempt any entity from measurement. We do not have either direct access to the data or a copy of all the iterative analyses at this time or we would include more specific data to demonstrate these findings. The analyses were in hand and were incorporated into our decision-making at the time that we developed the specifications.

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

See section above.

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

3. Feasibility

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

3a. Byproduct of Care Processes

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

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

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

If other:

3b. Electronic Sources

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

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

Some data elements are in defined fields in electronic sources

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

Detailed clinical data are needed. There are no technical barriers to capturing the necessary data in defined electronic fields in electronic health records. We view NQF endorsement as a step to help us to initiate a conversation to consider such inclusion.

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

Attachment:

3c. Data Collection Strategy

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

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

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

We have learned that chart review is a reliable and accepted method of measuring appropriate use. There are no technical barriers to incorporating structured fields to help assess the appropriateness of the visits in conjunction with the criteria outlined above and implemented in this measure, although such fields do not currently exist. We further demonstrated that our measure was able to identify differences in the proportion appropriate, such as those associated with age and race. For example, the overall level of appropriateness for children aged 2-5 was 54%, for children aged 6-1 was 44%, and for adolescents between 12 and 18, 48%. Because of these differences we have chosen to present the measure stratified by age group. We found that use of a clinical database was an inefficient way to identify eligible charts and thus have adapted eligibility criteria that rely on administrative data. Because chart review is relatively time consuming, we have articulated the specifications in a way that

represents a hybrid whereby administrative data can qualify a proportion of numerator events without chart review. Our paper data collection tool underwent a number of revisions for time and data collection efficiency and the chart review team demonstrated excellent agreement in data collection with a group kappa of .87 in identifying numerator events. Although the chart collection tool is not a formal part of this measure, we would be happy sharing a general version (data collection template) of it upon request.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None at present.

4. Usability and Use

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

4a. Accountability and Transparency

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

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

Specific Plan for Use	Current Use (for current use provide URL)
Not in use	

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

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

N/A

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
We are awaiting NQF endorsement for use. There are no policies or actions of the developer/steward or accountable entities that would restrict access to performance results of impeded implementation. Some potential users are awaiting NQF endorsement.

The topic of ED asthma overuse was assigned to our measure development project in the Pediatric Quality Measures Program by CMS, by far the largest single third party payer for medical care for children in the US, and by AHRQ. Major federal policy makers have indicated to us that these measures are a priority. This measure has received the imprimatur of the American Academy of Pediatrics as one of its high priority measures that emerged from their joint (with the American Board of Pediatrics) Measurement Alignment and Strategic Selection Work Group.

We have begun a dialogue with the CDC to consider use of this measure to serve their interests.

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

The CAPQuaM team includes multiple stakeholders, including payers and state agencies. Several are interested in using this measure and are awaiting NQF endorsement. As a part of our CAPQuaM work we will disseminate and assist in the implementation of this measure subsequent to endorsement. This measure has been approved for inclusion in the National Quality Measures Clearinghouse.

As noted above, the topic of ED asthma overuse was assigned to our measure development project in the Pediatric Quality Measures Program by CMS, by far the largest single third party payer for medical care for children in the US, and by AHRQ. Major federal policy makers have indicated to us that these measures are a priority. This measure has received the imprimatur of the American Academy of Pediatrics as one of its high priority measures that emerged from their joint (with the American Board of Pediatrics) Measurement Alignment and Strategic Selection Work Group.

We have begun a dialogue with the CDC to consider use of this measure to serve their interests.

Improvement

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

4b. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
Not in longitudinal use. As noted above, both high and low levels of appropriateness are interpretable and actionable as outcomes of asthma management. This measure of process provides information regarding the outcomes of asthma care – both access to care and quality of management. Its interpretation is synergistic with the CAPQuaM rate of asthma ED visit measure also developed in the PQMP and currently under review at NQF.

4c. Unintended Consequences

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

4c.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**
None observed.

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

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
The CAPQuaM team includes payers, MCO´s, state health programs, consumers, Accreditors, family advocates, clinicians, hospitals, and others and all have had the opportunity to participate in the dialogue that led to the measure development and to the interpretation of its findings

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.

Steering committee meetings, conference calls, email

4d2.2. **Summarize the feedback obtained from those being measured.**
We don´t distinguish by source of feedback, please see 4d2.3

4d2.3. **Summarize the feedback obtained from other users**
This measure has been received enthusiastically by our stakeholder partners. We highlight feedback in an area for which we received comment in the prior review of this measure.

This measure is unusual in that it offers value and opportunity for improvement regardless of results. Low levels of appropriateness suggest inefficiencies in primary care, as children who probably do not need to be in the ED because of the severity of their disease are nonetheless being brought to the ED by their caregivers. This may represent failures of asthma education or of availability/access/attractiveness of primary care in the context of acute concerns. Low levels of appropriateness also suggest that whatever measures are being used to assess how frequently children are using the ED are overestimating the clinical failure rate of asthma care, since the ED visits that are presumed to be clinical failures may represent something else.

Conversely, high levels of appropriateness suggest that the ED use for asthma is predominately among those children who are experiencing breakthrough in their clinical asthma. If the overall rate of use is high, this would raise concerns about the effectiveness of the asthma management in ambulatory care. But it does not point specifically to availability/access/attractiveness of primary care as does low rates of effectiveness.

Because of this bi-directionality of the measures interpretation, there is lower risk for gaming than with some other measures.

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
Feedback incorporated directly into the development process.

5. Comparison to Related or Competing Measures

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

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

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

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

5a. Harmonization of Related Measures
The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

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

5b. Competing Measures

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

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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Appendix_Asthma_5_12_12_16.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** University Hospitals Cleveland Medical Center
Co.2 **Point of Contact:** Lawrence, Kleinman, drlarrykleinman@gmail.com, 216-286-6969-
Co.3 **Measure Developer if different from Measure Steward:** University Hospitals Cleveland Medical Center
Co.4 **Point of Contact:** Lawrence, Kleinman, drlarrykleinman@gmail.com, 216-286-6969-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Role: Expert Panelists
Elizabeth Allen, M.D Ohio State University
Chitra Dinakar, M.D. Department of Pediatrics University of Missouri-Kansas City
Stephen Teach, M.D., M.P.H. Children's National Medical Center
Charles Macias M.D., M.P.H. Pediatrics, Baylor Texas Children's Hospital
Michael Cabana M.D., M.P.H. University of California, San Francisco

Barbara Yawn M.D., M.S. Olmstead Medical Center, University of Minnesota
Joan Connell M.D. University of North Dakota
Delaney Gracy M.D., M.P.H. The Children's Health Fund and Montefiore Children's Hospital
Sharlene Miner M.D. Emergency Medicine Inter Mountain Medical Center

ROLE: Steering Committee and Investigator
Wilson Pace, MD American Academy of Family Physicians – DARTNET Institute - University of Colorado
Lynn Olson, PhD American Academy of Pediatrics
Christina Bethell, PhD, MBA, MPH Child and Adolescent Health Measurement Initiative, Johns Hopkins University (Previous OHSU)
Elizabeth Howell, MD Icahn School of Medicine at Mount Sinai
Harold Kaplan, MD Icahn School of Medicine at Mount Sinai
Lawrence Kleinman, MD, MPH Icahn School of Medicine at Mount Sinai

Rebecca Anderson Mount Sinai Medical Center
Ira Nash, MD Northshore - Long Island Jewish Medical Center (previous Mount Sinai School of Medicine)
Eyal Shemesh, MD Icahn School of Medicine at Mount Sinai
Mary Barton, MD National Committee on Quality Assurance
Charles Homer, MD, MPH US Department of HHS (previous National Institute for Child Health Quality)
Marla Clayman, PhD American Institutes for Research (previous Northwestern University)
Foster Gesten, MD New York State Dept. of Health, Office of Health Insurance Programs
Jerod M. Loeb, PhD The Joint Commission
Robert Rehm National Committee for Quality Assurance
Steve Kairys, MD American Academy of Pediatrics/QuIIN
Erin DuPree, MD The Joint Commission (previous: Icahn School of Medicine at Mount Sinai)
David Baker The Joint Commission (Previous Northwestern)
Beverley Johnson, BSN* Institute for Patient- and Family-Centered Care
Doris Peter Consumers Union

ROLE: Senior Advisory Board Member and Investigator
Shoshanna Sofaer, DrPH American Institutes for Research (previous CUNY Baruch)
Lynne Richardson, MD Icahn School of Medicine at Mount Sinai
Ian Holzman, MD Icahn School of Medicine at Mount Sinai
Marilyn Kacica, MD New York State Dept. of Health, Division of Family Health

ROLE: Senior Advisory Board Member
Marc Lashley, MDAllied Pediatrics
Gary Mirkin, MD Allied Pediatrics
Barbara Kupferman, RN, CMCN, LNC AmeriChoice by United Healthcare
L. Gregory Pawlson, MD, MPH SE Healthcare QUALITY Consultants (previous Blue Cross Blue Shield)
John Santa, MD* (previous Consumers Union)
John Clarke, MD ECRI, PA Patient Safety Authority
Scott Breidbart, MD Empire Blue Cross Blue Shield/ Anthem
Jeff Terry, MBA GE Healthcare
Robert St. Peter, MD, MPH Kansas Health Institute
Ruth Stein, MD Montefiore Children's Hospital
Arthur Aufses, MD Icahn School of Medicine at Mount Sinai
Eric Rose, MD Icahn School of Medicine at Mount Sinai
Wendy Brennan, MS National Alliance on Mental Illness of New York City (NAMI - NYC Metro)
Laurel Pickering, MPH Northeast Business Group on Health
Tony Hope, BM BCh, PhD Oxford University, The Ethox Center
Martin Hatlie, JD Partnership 4 Pt. Safety, Consumers Advancing Patient Safety
Paul Wise, MD Stanford University
Mark Chassin, MD, MPH, MPP The Joint Commission
Ira Nash, MD Icahn School of Medicine at Mount Sinai
ROLE: Investigators/Key Staff
Rusty McLouth, MS AAFP
Kasey McCracken, MPH CAHMI
Elise Barrow, MPH Icahn School of Medicine at Mount Sinai
Kasey Coyne, BA Icahn School of Medicine at Mount Sinai
Natalia Egorova, PhD Icahn School of Medicine at Mount Sinai
Elizabeth Howell, MD Icahn School of Medicine at Mount Sinai
Harold Kaplan, MD Icahn School of Medicine at Mount Sinai
Melissa Manice, MPH Icahn School of Medicine at Mount Sinai
Joshua Oppenheimer, BA Icahn School of Medicine at Mount Sinai
Eve Oster, MD Pediatric Partners of Gwinnett (previous Icahn School of Medicine at Mount Sinai)
Daniel O'Sullivan, BS University of Virginia (previous Icahn School of Medicine at Mount Sinai)

Michael Parides, PhD	Icahn School of Medicine at Mount Sinai
Barbara Rabin, MHA	Icahn School of Medicine at Mount Sinai
Carolyn Rosen, MD	Icahn School of Medicine at Mount Sinai
Melissa Saperstein, MSW	Icahn School of Medicine at Mount Sinai
Eyal Shemesh, MD	Icahn School of Medicine at Mount Sinai
Andrew Ting, MD	Icahn School of Medicine at Mount Sinai
Adam Vella, MD	Icahn School of Medicine at Mount Sinai
Virginia Walther, MSW	Icahn School of Medicine at Mount Sinai
Jennifer Zeitlin, SD	Icahn School of Medicine at Mount Sinai, French Institute of Health and Medical Research (INSERM)
Marianne McPherson, PhD, MS	NICHQ
Joseph Anarella, MPH	NYS DOH, OHIP
Lee Sanders, MD, MPH	Stanford University
Kasey Coyne	Northwestern
Victoria Wagner	New York State Health Department Office Quality Patient Safety
Laura Soloway	New York State Health Department Office Quality Patient Safety
Kevin Malloy	New York State Health Department Office Quality Patient Safety
Judy Stribling	Icahn School of Medicine at Mount Sinai
Linda Radecki	AAP
Keri Thiessen	AAP – QuIIN
Laura Morris	New York State Department of Health
Olivia Lindly	OHSU
Kasey McCracken	OHSU
Sandeep Sharma	Icahn School of Medicine at Mount Sinai
Ann Nevar	UH Rainbow Babies and Children's Hospital/ CAPQuaM
Suzanne Lo	UH Rainbow Babies and Children's Hospital/ CAPQuaM
Amy Balbierz	Icahn School of Medicine at Mount Sinai
Samantha Raymond	Icahn School of Medicine at Mount Sinai
Allisyn Vachon	Icahn School of Medicine at Mount Sinai
Shannon Weber	Icahn School of Medicine at Mount Sinai
ROLE: Contractors	
Mary Fialkowski, MBA	The Joint Commission
Linda Hanold, MHSA	The Joint Commission
Elvira Ryan, MBA, BSN, RN	The Joint Commission
Sharon Sprenger, RHIA, CPHQ, MPA	The Joint Commission
Ann Watt, MBA, RHIA	The Joint Commission
Amy Slott	The Joint Commission
Denise Krusenoski	The Joint Commission
Elvira Ryan	The Joint Commission
Tasha Mearday	The Joint Commission

Measure Developer/Steward Updates and Ongoing Maintenance

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

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

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

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

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:**

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

Appendix

Figure 1

FIGURE 1

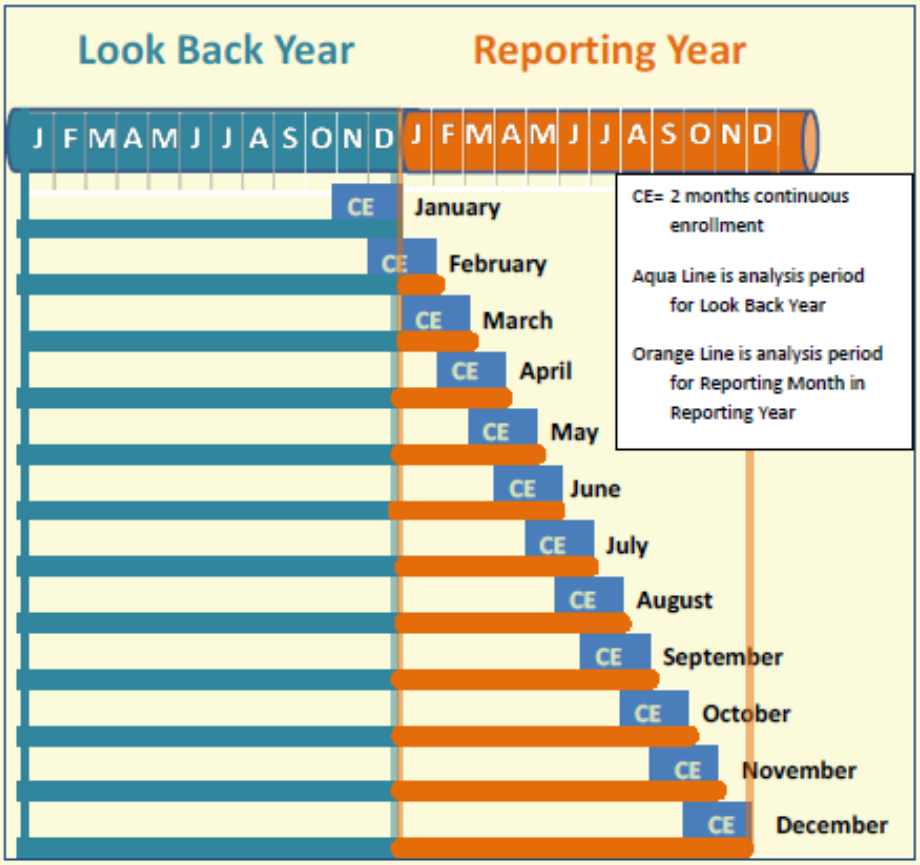


Diagram of Eligibility Requirements for Asthma Measure

Description of Diagram Components

Aqua: Look Back Year
Orange: Reporting Year
Blue: Continuous Enrollment (CE) period must be met to be eligible in a given reporting month
Black: Label of Reporting Month

Description of Eligibility Criteria Assessment

Eligibility Criteria are assessed month by month, using both the entire Look Back Year and the Reporting Year.

For any given reporting month, assess eligibility on 2 criteria. Eligible children are those that meet both of the following:

1) Has the child been continuously enrolled (CE) for the two months prior to the Reporting Month (3 months of continuous enrollment including the current Reporting Month)
2) Does the child meet the criteria for having been managed for persistent asthma in the assessment period (Aqua + Orange Lines)

Figure 2
Roadmap and overview of sample selection and measure calculation

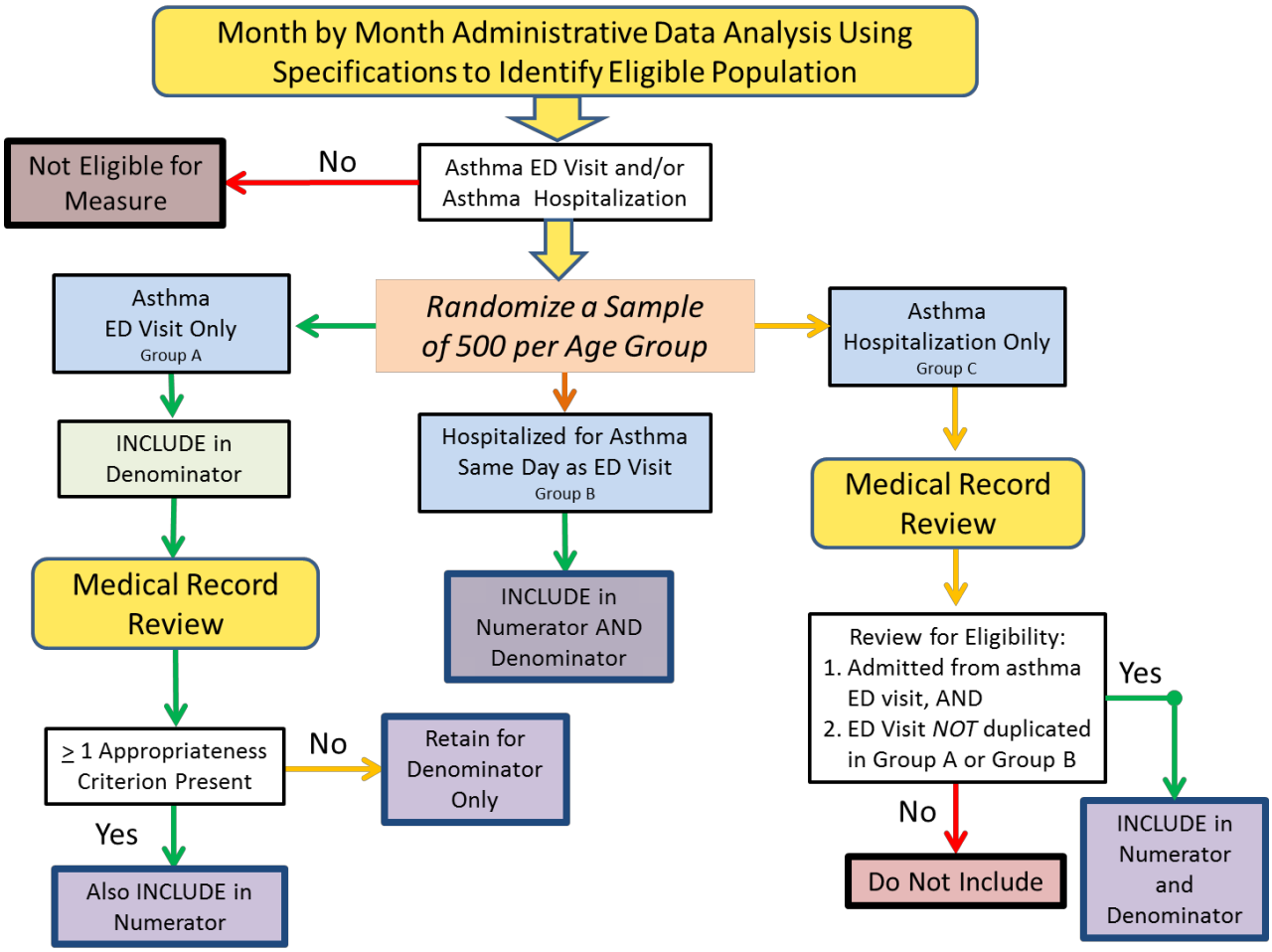


Figure 2 Notes:

- Determining eligibility for sample selection precedes determining eligibility for measure.
- On the basis of the Administrative Data Analysis, children who are potentially eligible for the measure will be identified and segregated into Groups A, B, and C (the blue boxes above).
- Children are eligible for Group B if three things are found in the administrative data: ED Visit; Hospitalization on same day and same institution; and Hospital discharge is after date of ED visit.
- National and NY State data suggest that approximately ¾ of childhood asthma hospitalizations are admitted from ED, that about 1 in 9 childhood asthma ED visits result in hospitalization and that children admitted from the ED may not have their ED visit coded in administrative data.
- Medical record review determines eligibility for numerator among the Group A children, all of whom have already qualified to be included in the denominator.
- Group B children are eligible for both the numerator and the denominator on the basis of administrative data analysis alone and do not require chart review.
- Medical record review determines eligibility for inclusion in the measure (denominator!) for Group C children. If they are eligible for the denominator (i.e. that have been admitted directly from an unduplicated ED visit) then they are also qualified for the numerator.

[i] The *k* value indicates a near perfect agreement (*k*: 0.81-1.0 between coded data and chart review data), and substantial agreement (*k*: 0.61-0.80).

[ii] We note that 1799 and 1800 are not directly applicable because they were tested at the score level. However, the scores were dependent upon definitions which use the same data element level as our measure and thus provide indirect evidence of the capacity of a measure using such data elements to produce valid scores.

Thus we cite them not as specific evidence of our score level performance of the submitted measure, but as evidence that the HEDIS measures that rely on the same administrative data elements for their denominator have the capacity to distinguish signal to noise at a very high level. While the evidence is indirect it is dispositive. That is, we assert that had the data elements been inadequate it would result in non-differential misclassification error which is a major bias towards the null thus introducing noise and reducing signal. That this does not happen to an appreciable degree specifically implies that the data elements function well – indeed this could be one rationale for why NQF allows the use of performance score level analysis in the first place. These findings provide strong indirect evidence of the validity of our approach to capturing the measure's denominator.

There is nearly complete overlap of the denominator codes and there is overlap of the denominator elements. Where codes differ it is specific to decisions made by the CAPQuaM expert panel which was aware of the NCQA measures. Review of the medication lists for 0036 reveal that all medication used by the submitted CAPQuaM measure are also in the HEDIS measure. The CAPQuaM measure excludes specifically short acting beta agonists and leukotriene inhibitors at the specific direction of the CAPQuaM expert panel. We also specify exclude indacaterol from the list of "asthma specific medications" since it is a long acting beta agonist which is only indicated in the USA for treatment of COPD, which is a specific exclusion criterion for this measure.

Further, we identify asthma visits and medications using the same data that an insurance company or Medicaid would use for payment, including ICD codes, CPT codes, and revenue codes. We have had conversations with expert coders and New York State Department of Health Office of Health Insurance Programs to confirm our choices. Our literature review found that while there is moderate agreement (kappa = 0.45 - 0.50) when comparing administrative data regarding the presence of constructs such as recent asthma attacks, use of asthma medications, attack or medication, attack and medication, using 1 year of administrative claims data to parent report, the agreement improves from 0.55 to 0.60 when using two years of data.(1) We expect that these kappas would be significantly higher were the analyses restricted to children with disease that met our construct criteria for identifiable asthma.

The literature further supports our work as highlighted above in the table and in more detail in our testing form 2b2.3 (validity testing).

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3136
Corresponding Measures:
De.2. **Measure Title:** GAPPS: Rate of preventable adverse events per 1,000 patient-days among pediatric inpatients
Co.1.1. **Measure Steward:** Center of Excellence for Pediatric Quality Measurement
De.3. **Brief Description of Measure:** GAPPS is a measure of the number of preventable adverse events per 1,000 patient-days among pediatric inpatients. It is designed to compare rates across institutions and over time. The GAPPS measure utilizes the GAPPS trigger tool to identify adverse events.
1b.1. **Developer Rationale:** Patient safety is a core domain of healthcare quality and a major focus for quality improvement efforts.(1,2) GAPPS is the first and only available global patient safety measure tailored for pediatric populations. By measuring preventable adverse event rates (i.e., harm) in inpatient pediatric populations, it provides important information to providers, hospital quality teams, and state health departments about outcomes of their patient care. Use of the measure will benefit patients, families, and providers because it enables stakeholders to identify and target areas of patient care that may benefit from quality improvement initiatives. Since GAPPS focuses on preventable adverse events, hospitals are able to assess and prioritize clinical areas with potential for immediate improvement.

The GAPPS Measure represents an opportunity to fill a notable void in safety measurement, specifically in pediatric preventable adverse events. National progress on quantifying and tracking AEs has stagnated because of the absence of an accepted national standard in all but a few defined areas. The GAPPS Measure can represent that national standard for pediatrics, and can be an important measure to evaluate and improve adverse event rates across different sites and time periods.

IMPORTANCE OF MEASURING ADVERSE EVENTS
Studies show that there is a high prevalence of medical errors and/or adverse events among the patient population. For example, in 1999, the Institute of Medicine (IOM) estimated that medical errors contribute up to 98,000 deaths and one million injuries each year.(2) In 2010, the Department of Health and Human Services' Office of the Inspector General estimated that 180,000 deaths due partly to adverse events occur among Medicare patients annually, making adverse events the third leading cause of death in the United States after heart disease and cancer.(3,4) Hospitalized pediatric patients, who tend to have unique diseases and care distinct from adult patients, are also vulnerable to high adverse event rates: published studies report 11.1 adverse drug events per 100 inpatient pediatric patients, 74 adverse events per 100 neonatal intensive care unit (NICU) patients, and 203 adverse events per 100 pediatric intensive care unit (PICU) patients.(5–9) Consequently, tracking adverse events in hospital settings is an important step towards understanding the current state of clinical care and creating initiatives aimed towards improving clinical quality.

Measuring preventable adverse event rates may also help hospitals better understand different aspects of their clinical quality. Some studies have found correlations between patient harm and other quality aspects such as performance on clinical processes of care and other health outcomes.(10–14) For instance, studies suggest that patients who experience healthcare-related harms have greater odds of in-hospital and 30-day mortality, as well as 30-day readmission.(11–14) As such, measuring preventable adverse event rates is an essential first step for hospitals to understand and improve their patient care.
IDENTIFICATION OF INPATIENT ADVERSE EVENTS
Various approaches exist for identifying adverse events. Voluntary passive reporting systems are commonly employed but recognized to have low sensitivity.(15) A more reliable, sensitive methodology for capturing data on the safety of hospital care is thus essential.(16,17)

In 2003, AHRQ released its Patient Safety Indicators (PSI), developed in response to a congressional mandate to reduce medical errors.(18) PSIs are intended to identify events that most likely resulted from preventable medical errors.(18–20) PSIs have been used with some success but have a number of limitations, in part due to their reliance on administrative data. They have also been found to have low sensitivity.(17,18)

Use of trigger tools has been shown to be a faster, more sensitive, and more reliable method of adverse event detection than other approaches.(5,9,17,21–23) "Triggers" are red flags in a medical record that may indicate the presence of an underlying adverse event and prompt further inspection to determine whether an adverse event occurred.(21,22) An example trigger is the documented administration of an antidote-type medication (e.g., naloxone). Once a trigger is found, an in-depth review is undertaken to determine whether an adverse event occurred. In the case of naloxone, administration may indicate an adverse event occurred if the drug was given to counteract on overdose of opioids given in the hospital but may not if the overdose occurred due to voluntary recreational opioid use. Trigger tools detect adverse events in a high percentage of hospitalizations, ranging in published reports from 19% to 63%, and have evolved significantly over time.10-12

The Global Trigger Tool for Measuring Adverse Events (GTT), developed by the Institute for Healthcare Improvement (IHI), has become widely accepted as an effective approach for identifying adverse events in hospitalized adult patients.(2,9,15,23–26) The GTT approach identifies 10 times more adverse events than AHRQ's PSIs and almost 100 times more events than voluntary reporting.(1,17) However, the GTT has an exclusion of patients under age 18 so does not work for a pediatric population.

PEDIATRIC INPATIENT PATIENT SAFETY: LACK OF STANDARDIZED QUALITY MEASUREMENT

Although one study determined that a version of the GTT applied to the pediatric population could identify pediatric adverse events, the authors and other experts called for development of a standardized pediatric tool that focuses specifically on the problems of hospitalized children and that encompasses the breadth of inpatient pediatric care.(12, 18, 25, 29) The absence of a comprehensive pediatric trigger tool is a recognized limitation in quantifying the full scope of pediatric adverse events. An early effort to develop a pediatric-focused trigger tool led to the development of the Canadian Pediatric Trigger Tool.(30, 31)

We developed GAPPS to meet the need for a comprehensive, sensitive measure of pediatric patient safety. Our focus was on developing a global trigger tool for pediatric patients that could be more reliably applied across different hospital sites, both academic and community, than previous efforts. In addition, we sought to further refine the list of triggers to make a more robust global trigger list. We used methods similar to those used for GTT, including review of published tools and manual medical record review by experts in patient safety, which has been demonstrated to be a crucial component of developing patient safety measures.(30) We also utilized the RAND/UCLA Appropriateness Method, a 16-center field study, and post-analysis refinement of the trigger list to ensure GAPPS includes a more comprehensive trigger list than previous trigger tools.

GAPPS offers an enhancement in trigger tool methodology in that, unlike GTT, it requires that reviewers assess preventability. In the five years since IHI released the second edition of GTT, patient safety experts and national fiscal and quality improvement policies have increasingly focused on addressing preventable adverse events. The GAPPS measure uses the same approach to rate preventability as the North Carolina Patient Safety study, which was found to ascertain preventability with a high degree of reliability.(31)

DISPARITIES IN RISK OF HARM

Children with special healthcare needs experience elevated rates of medical errors.(32) Among hospitalized pediatric patients, those with chronic conditions are at significantly higher risk for medical errors than those without chronic conditions.(32)

What is known about racial/ethnic disparities in patient safety, particularly among children, is limited.(33) Black and Hispanic newborns are at higher risk of birth trauma.(18) In addition, extrapolations from associations between race/ethnicity and known risk factors for harm suggest that Black and Hispanic children are likely at greater risk of harm than White children. Because severity and complexity of illness increase the risk of errors, and Black and Hispanic children are at higher risk for more complex conditions, these children are at greater risk for adverse events.(33) Emergency room visits are also associated with increased risk of adverse events, and Black and Hispanic children are known to visit the emergency department more frequently.(28,33)

COSTS OF GAPPS UPTAKE

As is the case with all new quality measures, we recognize that there may be concerns about the uptake of GAPPS due to hospital resource constraints. AE monitoring, like patient experience surveying, is not possible using administrative data and therefore requires additional resources. As previously discussed, intensive measurement of preventable AEs is one of the most important areas of quality measurement, one that has to this point been inadequately addressed in pediatric patient populations. The

capacity of the GAPPS Measure to spur nationwide improvement in pediatric patient safety represents a significant potential return on investment, one that outweighs implementation resource concerns.

It is important to note that some of the resources and infrastructure to successfully implement the GAPPS measure are already in place, as states commit to systems of mandatory reporting of certain adverse events. About half of states currently have mandatory reporting systems in place, with many others reporting AEs on a voluntary basis.(34) GAPPS will require an augmentation of processes hospitals are already initiating to report AEs, but will yield far greater returns. Far from representing a duplicative burden, the GAPPS Measure can take advantage of recent prioritization of patient safety and the resources hospitals are putting in place to evaluate quality in this domain. The marginal resources used for GAPPS implementation provide a drastically more robust assessment of hospital safety than currently in place, as global trigger tools have been shown to capture up to ten times more AEs than alternative AE measurement methods.(17,35)

REFERENCES

1. Levinson DR. Adverse Events in Hospitals: Methods for Identifying Events. Washington, DC: Office of the Inspector General, Department of Health and Human Services; 2010. Report No.: OEI-06-08-00221.
2. Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human: Building a Safer Health System. Washington, DC: Institute of Medicine Committee on the Quality of Health Care in America; 1999.
3. James JT. A new, evidence-based estimate of patient harms associated with hospital care. J Patient Saf. 2013 Sep;9(3):122–8.
4. Allen M, ProPublica. How Many Die From Medical Mistakes In U.S. Hospitals? [Internet]. NPR.org. [cited 2014 Jun 14]. Available from: http://www.npr.org/blogs/health/2013/09/20/224507654/how-many-die-from-medical-mistakes-in-u-s-hospitals
5. Sharek PJ, Classen D. The incidence of adverse events and medical error in pediatrics. Pediatr Clin North Am. 2006 Dec;53(6):1067–77.
6. Pediatric Patient Safety Measurement Tool (PPSMT) Pilot Toolkit. Children's Hospital Association; 2012.
7. Stockwell DC, Slonim AD. Quality and safety in the intensive care unit. J Intensive Care Med. 2006 Aug;21(4):199–210.
8. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. JAMA J Am Med Assoc. 1997 Jan 22;277(4):307–11.
9. Takata GS, Mason W, Taketomo C, Logsdon T, Sharek PJ. Development, testing, and findings of a pediatric-focused trigger tool to identify medication-related harm in US children's hospitals. Pediatrics. 2008 Apr;121(4):e927-935.
10. Calder L, Tierney S, Jiang Y, Gagné A, Gee A, Hobden E, et al. Patient safety analysis of the ED care of patients with heart failure and COPD exacerbations: a multicenter prospective cohort study. Am J Emerg Med. 2014 Jan;32(1):29–35.
11. Joice GA, Deibert CM, Kates M, Spencer BA, McKiernan JM. "Never events": Centers for Medicare and Medicaid Services complications after radical cystectomy. Urology. 2013 Mar;81(3):527–32.
12. Rosen AK, Geraci JM, Ash AS, McNiff KJ, Moskowitz MA. Postoperative adverse events of common surgical procedures in the Medicare population. Med Care. 1992 Sep;30(9):753–65.
13. Rosen AK, Loveland S, Shin M, Shwartz M, Hanchate A, Chen Q, et al. Examining the impact of the AHRQ Patient Safety Indicators (PSIs) on the Veterans Health Administration: the case of readmissions. Med Care. 2013 Jan;51(1):37–44.
14. Sukumar S, Roghmann F, Trinh VQ, Sammon JD, Gervais M-K, Tan H-J, et al. National trends in hospital-acquired preventable adverse events after major cancer surgery in the USA. BMJ Open. 2013;3(6).
15. Sharek PJ, Horbar JD, Mason W, Bisarya H, Thurm CW, Suresh G, et al. Adverse events in the neonatal intensive care unit: development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. Pediatrics. 2006 Oct;118(4):1332–40.
16. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. N Engl J Med. 2010 Nov 25;363(22):2124–34.
17. Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. "Global Trigger Tool" Shows That Adverse Events In Hospitals May Be Ten Times Greater Than Previously Measured. Health Aff (Millwood). 2011 Apr 1;30(4):581–9.
18. Miller MR, Elixhauser A, Zhan C. Patient Safety Events During Pediatric Hospitalizations. Pediatrics. 2003 Jun 1;111(6):1358–66.
19. Dynan L, Goudie A, Smith RB, Fairbrother G, Simpson LA. Differences in Quality of Care Among Non–Safety-Net, Safety-Net, and Children's Hospitals. Pediatrics. 2013 Feb 1;131(2):304–11.
20. Miller MR, Zhan C. Pediatric Patient Safety in Hospitals: A National Picture in 2000. Pediatrics. 2004 Jun 1;113(6):1741–6.
21. Rozich JD, Haraden CR, Resar RK. Adverse drug event trigger tool: a practical methodology for measuring medication related harm. Qual Saf Health Care. 2003 Jun 1;12(3):194–200.
22. Resar RK, Rozich JD, Classen D. Methodology and rationale for the measurement of harm with trigger tools. Qual Saf Health Care. 2003 Dec 1;12(suppl 2):ii39-ii45.

23. Agarwal S, Classen D, Larsen G, Tofil NM, Hayes LW, Sullivan JE, et al. Prevalence of adverse events in pediatric intensive care units in the United States. Pediatr Crit Care Med J Soc Crit Care Med World Fed Pediatr Intensive Crit Care Soc. 2010 Sep;11(5):568–78.
24. Thomas EJ, Studdert DM, Burstin HR, Orav EJ, Zeena T, Williams EJ, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Med Care. 2000 Mar;38(3):261–71.
25. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. CMAJ Can Med Assoc J J Assoc Medicale Can. 2004 May 25;170(11):1678–86.
26. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The Quality in Australian Health Care Study. Med J Aust. 1995 Nov 6;163(9):458–71.
27. Kirkendall ES, Kloppenborg E, Papp J, White D, Frese C, Hacker D, et al. Measuring adverse events and levels of harm in pediatric inpatients with the Global Trigger Tool. Pediatrics. 2012 Nov;130(5):e1206-1214.
28. Matlow AG, Baker GR, Flintoft V, Cochrane D, Coffey M, Cohen E, et al. Adverse events among children in Canadian hospitals: the Canadian Paediatric Adverse Events Study. CMAJ Can Med Assoc J J Assoc Medicale Can. 2012 Sep 18;184(13):E709-718.
29. Stockwell D, Bisarya H, Classen D, Kirkendall E, Landrigan C, Lemon V, et al. A trigger tool to detect harm in pediatric inpatient settings. Pediatrics. 2015;
30. Scanlon MC, Harris JM, Levy F, Sedman A. Evaluation of the Agency for Healthcare Research and Quality Pediatric Quality Indicators. Pediatrics. 2008 Jun 1;121(6):e1723–31.
31. Sharek PJ, Parry G, Goldmann D, Bones K, Hackbarth A, Resar R, et al. Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. Health Serv Res. 2011 Apr;46(2):654–78.
32. Ahuja N, Zhao W, Xiang H. Medical Errors in US Pediatric Inpatients With Chronic Conditions. Pediatrics. 2012 Oct 1;130(4):e786–93.
33. Flores G, Ngui E. Racial/Ethnic Disparities and Patient Safety. Pediatr Clin North Am. 2006 Dec;53(6):1197–215.
34. QuPS.org - Medical Errors and Patient Safety - US States Comparison - Overview - [Internet]. [cited 2016 Nov 16]. Available from: http://www.qups.org/med_errors.php?c=internal&id=195
35. Office of the Inspector General. Adverse Events In Hospitals: National Incidence Among Medicare Beneficiaries. Department of Health and Human Services,; 2010.

Numerator Statement: The number of preventable adverse events found in a patient sample.
Denominator Statement: The denominator is 1,000 patient-days for all sampled pediatric patients who meet inclusion, but not exclusion, criteria.
Denominator Exclusions: N/A

Measure Type: Outcome
Data Source: Electronic Health Record (Only), Paper Records
Level of Analysis: Facility

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary

- This is an outcome measure. The developer provides a conceptual logic model linking the identification of preventable adverse events (AEs) with improved patient safety.
- In addition to the rationale, the developer also cited 32 studies and articles to support the measure:
  - The use of trigger tools helps identify cases of patient harm.
  - A relationship between patient safety and other aspects of quality exists:

        - Several studies show patients who experience health-care related harms have greater odds of in-hospital, 30-day mortality, and 30-day readmission.
        - While the research is primarily in adults, the developer asserts that "It is likely that the same general association hold true for pediatric populations".
    - Measuring patient safety can drive quality improvement.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Guidance from the Evidence Algorithm: Patient-reported outcome (Box 1)→ Relationship between PRO and provider action (Box 2)→ Pass

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

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

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

- The developer states GAPPS is the first and only available global patient safety measure tailored for pediatric populations, and it notes that the current comparable tool, the Global Trigger Tool (GTT), does not assess adverse events in the pediatric population. Specifically, "the GTT has an exclusion of patients under age 18 so does not work for a pediatric population… Hospitalized pediatric patients, who tend to have unique diseases and care distinct from adult patients, are also vulnerable to high adverse event rates."
- The developer also states the GAPPS tool requires that reviewers assess preventability, while the GTT does not.
- As this is a new measure, no widespread gap information is available. The developer reports, however:
    - Literature demonstrates that there are 11.1 adverse drug events per 100 inpatient pediatric patients, 74 adverse events per 100 neonatal intensive care unit (NICU) patients, and 203 adverse events (AE) per 100 pediatric intensive care unit (PICU) patients.
    - During its field test of 16 hospitals from across the United States. 3,790 hospitalizations occurring between 2007-2012 were included in the analysis.
        - 414 AEs identified, 210 (50.7%) of which were preventable.
        - This represents 9.5 preventable AEs [CI 8.2-10.8]/1,000 patient days.
        - Compared to community hospitals, academic hospitals had higher preventable harm rates (13.1 [CI 11.4-15.2] vs. 2.4 [CI 1.5-3.8] AEs/1,000 patient days, p<0.001).

Disparities

- The developer cited data extrapolated from testing and report the following for patients based on race/ethnicity, number of chronic conditions, and insurance statuses.

Table 1 –Preventable AEs per 1,000 patient days by race/ethnicity (n = 3,231)			
Race/Ethnicity	n (%)	prev AE rate	P-Value
White	2,152 (56.8%)	8.9	Reference
Black	476 (12.6%)	5.9	0.13
Hispanic	419 (11.1%)	15.9	0.002
Other	184 (4.9%)	11.1	0.047

Table 2 – Preventable AEs per 1,000 patient days by numbers of Chronic Condition Indicators per patient (n= 3,524)			
Chronic Condition Indicators	n (%)	prev AE rate	P-Value
0 body system	1,990 (52.5%)	6.5	Reference
1 body system	1,085 (28.6%)	9.5	0.04

2 body systems	321 (8.5%)	17.9	<0.001
3 or more body systems	128 (3.4%)	19.8	<0.001

Table 3 – Preventable AEs per 1,000 patient days by insurance type (n = 3,468)			
Insurance	n (%)	prev AE rate	P-Value
Public	1,300 (37.5%)	12.1	Reference
Private	2,064 (59.5%)	8.5	0.02
No Insurance	104 (3.0%)	3.9	0.11

Question for the Committee:

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

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

Committee pre-evaluation comments

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

- The literature review and resulting adverse events was concerning, particularly with higher pediatric than adult rates, and disparities regarding CSHCN, public insurance, and African-American/Hispanic children. The most interesting aspect is the innovation regarding prevention. It is agreed that the measure passes.
- The use of trigger lists to more efficiently identify hospital-based adverse events than chart review alone is evident in the adult literature, but not in the pediatric literature. However, inpatient hospital care in adults parallels that of children in that the processes followed and systems are often similar and can be modified in a way to prevent adverse events and medical errors. That said children are not small adults and the reasons that they are hospitalized are different from adults. As well their small size gives treatments like fluids and medication doses especially those with known toxicities like kidney and liver less margin of error. Identifying types of adverse events and the most common ones that are preventable logically leads to the improving the processes that allow the errors or events to occur and prevent the recurrence. This is a well-established quality improvement approach to sentinel events, identification of the root cause, and modification of the process. One question would be whether or not the chosen triggers are as applicable to smaller hospitals without pediatric or neonatal ICU care as they are to larger, referral or pediatric specialty hospitals.
- Again, this measure also assesses prevention. It is concerning that almost all children in NICU will experience one adverse event and children in PICU average 2. Also concerning is that academic hospitals have more adverse events than community hospitals, yet they are supposed to specialize. Again, the concerning disparities exist for children with chronic conditions, race/ethnicity, and type of hospital and type of insurance. Rating: high
- The data presented shows that solely reported adverse events and medical errors undercounts the actual events. Chart reviews of hospitalizations can identify more adverse events, but are time-consuming and resource intensive (and costly) to perform. The adult quality system has shown that using a trigger list especially in conjunction with an electronic health record can pick up more adverse events that reporting with significantly fewer resources allowing hospital systems to approach improving the systems that lead to those adverse events where they are preventable. Because pediatric hospital care differs from adult care both because of the conditions for which children are admitted and for the treatment they are given and their relative treatment error risks related to size, having a comparable trigger list for pediatrics would allow system improvement to result in safer care for children. Racial and ethnic disparities in numbers of reported adverse events may be reflective of the hospitals in which they receive their care. The gap in care (or safety) is really the difference between the rate of reported adverse events and those that would be captured by using a trigger list.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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

Data source(s): Electronic Health Record (Only), Paper Records

Specifications:

- Level of analysis: facility/hospital
- Interpretation of score: Better quality = lower score
- Numerator: *the number of preventable adverse events found in a patient sample.*
    - Adverse events are defined as "unintended physical injuries resulting from or contributed to by medical care that require additional monitoring, treatments, or hospitalizations, or that result in death."
    - The GAPPS measure requires that two physicians analyze and independently rate the preventability of each AE case under review. When disagreement exists, the reviewers discuss the rationale for their ratings until both agree on whether the AE is preventable. A third physician is consulted in the rare occasion the two physicians continue to disagree.
    - The list of triggers differs between the Automated and Manual tools.
- Denominator: *The denominator is 1,000 patient-days for all sampled pediatric patients who meet inclusion, but not exclusion, criteria*. It includes all patients who meet the following:
    1. Patients <18 years of age at admission;
    2. Patients with length of stay greater than or equal to 24 hours;
    3. Patients admitted for acute care. Acute care does not include patients discharged from the Emergency Department (ED) without admission to the hospital, or patients in rehabilitation and residential units, non-acute inpatient psychiatric units, newborn nurseries, and day treatment areas. If a patient is initially admitted acutely, but subsequently transferred to inpatient psychiatric care, the acute portion of the hospitalization should be included; and
    4. Patients who were discharged from, who were transferred out of, or who died during the inpatient or observation hospital stay.
- Exclusions: There are no exclusions, but the denominator definition notes that patients with inpatient LOS <24 hours are not included because patients with brief hospital stays are less likely to have received the amount of medical intervention necessary to evaluate the quality of care.
- A calculation algorithm for the measure is provided.
- Sampling instructions are provided.
- A risk adjustment statistical model and stratification analysis is provided.
- The GAPPS measurement tool is provided.
- A data dictionary containing a list of automated triggers, examples of programming for automated triggers, and ICD-9/ICD-10 codes is included.
- The developer notes that while the triggers differ in the automated vs. manual lists, this is an initial step to speed implementation and reduce human errors during initial review; implementation to identify adverse events remains that same. The developer posits there should be no obvious difference in the AE identification step (i.e., performance score) because this remains the same (i.e., two physicians comparing).

Questions for the Committee:

- *Is the list of automated triggers (more comprehensive) appropriate, and*
- *Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Do you agree with the developer's conclusion that it is likely this measure can be consistently implemented?*

2a2. Reliability Testing, Testing attachment

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

SUMMARY OF TESTING

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

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

- The developer used the data of hospitalizations from 2007-2012 at 16 hospital sites (4 hospitals in each of the four U.S. Census regions; 8 teaching, 8 non-teaching).

Total of 3,814 medical records reviewed across the 16 hospital sites (≈240 records/hospital).

Method(s) of reliability testing

- The developer provided a flowchart of the medical review process in the GAPPS study; 10% of the original sample (i.e., ~24 charts) were further reviewed.
- The developer used a Kappa statistic for variables with only two possible outcomes and a weighted Kappa computed with Fleiss-Cohen weights for variables with more than two possible ordinal outcomes.
- Categorization of Landis and Koch is used to interpret reliability for ranges of Kappa scores:
  - k <0: poor, k = 0.00-0.20: slight, k = 0.21-0.40: fair, k = 0.41-0.60: moderate, k = 0.61-0.80: substantial, k = 0.81-1.00: almost perfect

Results of reliability testing

- The developer reported both primary reviewers agreed on:
  - The total number of AEs 88% of the time for a particular hospitalization.
  - Whether a record did or did not contain at least one AE 92% of the time for a particular hospitalization.
  - Whether they identified the same AEs 62% of the time for a particular hospitalization.
- According to the developer, the reliability for internal primary reviewers (n = 379) was "substantial" for:

Table 3 - Internal Primary Reviewer (Nurse) Reliability:

Primary Review (Nurse)	Kappa	Agreement
# Triggers	0.68	83.1%
Any Trigger	0.67	87.9%
# Adverse Events	0.73	88.4%
Any Adverse Events	0.69	91.8%

- According to the developer, the reliability for internal secondary reviewers in Group A versus Group B (n = 379) was "substantial" for:

Table 4 - Internal Secondary Reviewer (Physician) Reliability:

Secondary review (Physician)	Kappa	Agreement
Verification of Same Adverse Events	0.81	91.6%
AE Severity	0.86	89.6%
AE Preventability	0.72	77.4%

- Both secondary reviewers agreed on:
  - The total number of AEs 92% of the time.
  - Whether a record did or did not contain at least one AE 94% of the time.
  - "In some cases, however, the AEs identified in the medical records differed."

Question for the Committee:

- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm: Empirical reliability testing conducted using statistical tests (Box 2) → Reliability testing conducted with computed performance measure scores (Box 3) → Appropriate method described (Box 5) → Moderate certainty (Box 6b) → Moderate

The highest eligible rating is MODERATE.

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

2b. Validity

2b1. Validity: Specifications

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

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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

SUMMARY OF TESTING

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

Method of validity testing of the measure score:

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

Validity testing method:

- The developer conducted empiric validity testing at the data element-level (identification of AE). The developer notes that no true "gold standard" exists as of now, so used the "general reviewer" as compared to trained external reviewers (the "gold standard").
- Face validity testing and the RAND/UCLA Appropriateness Method was used to develop the draft trigger tool. NQF's face validity assessment requires a specific assessment at the measure score level, which the developer does not report.

Validity testing results:

Table 5:

Identifying a record with 1 or more AEs	Specificity	Sensitivity
Internal primary reviewers versus External expert primary reviewers	0.91	0.40
Internal secondary reviewers versus External expert secondary reviewers	0.95	0.33

- The developer reports the sensitivity was 0.40 for primary reviewers and 0.33 after taking secondary reviewer verification into account.
- The developer posits that the lower sensitivity is likely due in part to the novice reviewers' lack of experience with the tool and their inability to make up for their inexperience by increasing the amount of time they took to perform their review, given that there was a 30-minute time limit per record.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*

- *Is the Committee concerned about the low sensitivity and/or does it wish to discuss with the developer the degree to which training (noted elsewhere) improves this value?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The developer states there are no exclusions and therefore an exclusion analysis was not applicable.
- The developer notes in the calculation algorithm: Patients with inpatient LOS <24 hours are excluded because patients with brief hospital stays are less likely to have received the amount of medical intervention necessary to evaluate the quality of care.
    - Patients =18 years of age at admission are excluded because the Center of Excellence for Pediatric Quality Measurement's (CEPQM) task was to create a tool for measuring patient safety in the pediatric age group (i.e., <18 years of age). With this in mind, GAPPS is designed to perform exclusively in pediatric patients.
- In addition, the denominator inclusion criteria require that patients are <18 years of age at admission; and that patients have a length of stay (LOS) greater than or equal to 24 hours. The rationale is that:
    - Patients with length of stay (LOS) =24 hours;
    - Patients admitted for acute care. Acute care does not include patients discharged from the ED without admission to the hospital; or patients in rehabilitation and residential units, non-acute inpatient psychiatric units, newborn nurseries, and day treatment areas. If a patient is initially admitted acutely but subsequently transferred to inpatient psychiatric care, the acute portion of the hospitalization should be included; and
    - Patients who were discharged from, who were transferred out of, or who died during the inpatient or observation hospital stay.

Questions for the Committee:

- *Is the lack of exclusions appropriate?*
- *Do you agree with the denominator definition, which does not include patients with LOS 24 hours because "patients with brief hospital stays are less likely to have received the amount of medical intervention necessary to evaluate the quality of care."?*

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

Conceptual rationale for SDS factors included? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- The developer evaluated patient age, sex, number of chronic conditions (based on AHRQ's Chronic Condition Indicator [CCI]) and service type to adjust for the differences in case-mix using a mixed effects negative binomial regression.
- Associated with PAE (preventable adverse event):
    - Age group (both in teaching and community hospitals)
    - Surgery services vs medical services (both in teaching and community hospitals)
    - Number of chronic conditions (patients with 2+ CCIs were significantly more likely to have PAEs compared to patients with 0 or 1 CCIs in both community and teaching hospitals)
- Not associated with PAE:
    - Gender (in either teaching or community hospitals)
- The developer does not adjust for SDS, noting the measure focuses on in-hospital processes of care that should be equally applied to all and are within the control of the healthcare system. The developer notes it recommends reporting of stratified analyses by SDS to facilitate identification of disparities in care.

Testing results

Table 6 - Model results for teaching and community hospitals:

	Teaching Hospitals		
	Coefficient	95% confidence interval	p-value
Number of CCIs (0 or 1=reference)			
2+ CCIs	0.410	0.052, 0.767	0.02
Age group (< 3 years = reference			
≥3 years and <10 years	0.458	0.016, 0.899	0.04
≥10 years and <18 years	0.439	0.056, 0.822	0.02
Service type (medical = reference)			
Surgery	0.638	0.225, 1.051	0.002
β_0	-4.724	-5.000, -4.448	<0.001

	Community Hospitals		
	Coefficient	95% confidence interval	p-value
Number of CCIs (0 or 1=reference)			
2+ CCIs	1.406	-0.055, 1.867	0.06
Age group (< 3 years = reference			
≥3 years and <10 years	1.450	-0.062, 2.962	0.06
≥10 years and <18 years	2.041	0.704, 3,378	0.003
Service type (medical = reference)			
Surgery	0.676	-0.652, 2.005	0.32
β_0	-7.060	-7.975, -6,145	<0.001

Based on the above models, the developer recommends stratification by hospital type. Additionally, it recommends use of the following three categorical variables in the GAPPS case-mix adjustment model: age, number of CCIs, and service type.

Questions for the Committee:

- *Is the risk-adjustment strategy included in the measure appropriate?*
- *Are the candidate and final variables included in the risk adjustment model adequately described for the measure to be implemented?*
- *SDS factors are not included in the risk adjustment model and the rationale provided. Does the Committee agree with this approach?*

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

- The developer tested the measure in 16 academic and community hospitals that care for children across the United States (discharges 2007-2012). The developer reports:
  - Harm due to medical care remained common and did not decrease significantly over the six year period in both community and academic centers

- The developer states, "effectively controlling pediatric patient safety problems has proven similarly difficult in both settings, despite their baseline differences in populations and harm epidemiology."
  - Approximately half of all harms are preventable
    - The range of preventable AEs was 0% to >20% per 1,000 hospital days
    - Most common AE's were on severity level of E and F
  - There are wide disparities in the rates of harm in academic and community hospitals
    - The developer states the reasons for the differences are unclear, but the "major differences in the frequency of complex chronic conditions as well as in the types and severity of illness seen in the two types of hospitals likely explain much of the difference."

- The developer includes several figures demonstrating the differences in performance:
  - Figure 3 - Distribution of preventable AE rates by hospital
  - Figure 4 - Severity of all harms and preventable harms
  - Figure 5 - Distribution of Adverse Events (AEs) by Hospital and Clinical Characteristics
  - Figure 6 – Rates of all harms, preventable harms, and high-severity harms per 1,000 patient-days, according to quarter

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- NA

2b7. Missing Data

- There were two occurrences of missing documents (medical records) for a selected hospitalization in study (n=3,814). In this case the record was removed from consideration and replaced by another hospitalization.
- The developer concluded that there is "no reason to believe missing data systematically biased performance results in any way."

Guidance from the Validity Algorithm: Specifications consistent with evidence (Box 1) → Potential threats to validity assessed (Box 2)→ Empirical validity testing conducted using the measures as specified and appropriate statistical test→ No validity testing with computed measure score (Box 6) → Validity testing with patient-level data elements (Box 10) → Method described and appropriate for all critical data elements (Box 11)→ Results indicate high or moderate certainty or confidence the data used in the measure are valid (Box 12a) → Moderate.

The highest eligible rating is MODERATE.

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

Committee pre-evaluation comments

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

- Although the data elements are clearly defined, it is concerning that those in the ER yet not admitted are not included as lack of treatment could lead to an adverse event. Note: the link for the data dictionary didn't work. The list of triggers is fairly comprehensive yet it is unclear whether UTIs from catheterization, which is fairly common, are included. The calculation algorithm is clear. it is agreed that this measure could be consistently implemented.
- The triggers represent events that put children at risk. While not all hospitals will have electronic records, the manual format might be able to use some of the medication triggers by identifying children that had been administered those medications as inpatients through the pharmacy system. A review of those charts could then occur. Listing the questions as "no" for the automated review leaves the impression that those are not able to be done using an automated system when in fact they were not used in the automated because in one system they couldn't be programmed into the record pull algorithm. The review triggers are clear as is the review process, sampling method, and calculations and this should be reproducible in a number of different hospital environments. The adult trigger list concept has been implemented successfully in adult hospitals.

- It is noted that that only the data element was used for reliability testing. It is also noted that there is substantial reliability for both primary and secondary reviewers. It is agreed that differences in performance can be identified. Rating: moderate.
- The reliability testing between the reviewers was substantial for the # of adverse events although the reviewers did not always identify the same adverse events and almost perfect for the level of severity and whether or not the adverse event was preventable. This method demonstrates that it picks up numbers of adverse events as well as severity and preventability at least with substantial reliability between reviewers.
- For 2b1 it is agreed that the specifications are consistent with the evidence. For 2b2, only the data element testing against a gold standard is utilized and it is noted no true gold standard exists. It is unclear if both face validity and empirical validity testing are used as clarification is need whether RAND is empirical validity testing. It is concerning that low sensitivity is attributed to lack of reviewer training. This measure, particularly the prevention piece, is an indicator of quality.
- The specificity or probability of detection is high (>0.9) for both the novice and expert reviewers although the sensitivity is low. Is there evidence that sensitivity improves as experience doing the reviews grows? Does missing some adverse events while capturing others impact on the utility of the tool? Does it still allow for quality and safety improvement? Did the authors identify whether or not there were systemic errors in the reviewers work in order to find a way to improve the sensitivity? It would be helpful to have some summary information about the hospitals and sites that tested this especially those that are not primarily children's hospitals including the numbers of pediatric admissions they have, whether or not they have a neonatal ICU and/or a pediatric ICU. This information might help clarify how applicable this would be to hospitals that are smaller and not primarily pediatric in patient population.
- For 2b3, it is concerning that nonadmission from the ER is excluded as stated previously. It is also unclear if this also includes mental health. For 2b4 it is reassuring that both statistical model and stratification are used but the acronym SDS needs to be defined. For 2b5, it is concerning that SDS factors are not included as disparities exist in these areas. It is also concerning that there was no improvement in the 6 year period and that half of adverse events are preventable. For 2b6 it is unclear why there is no comparability of data sources. For 2b7, it is agreed that the case of missing data was appropriately replaced and did not result in bias. Rating: moderate.
- The exclusions are appropriate to get a solely pediatric population with a hospital admission (rather than an observation day which might occur outside of the hospital ward and therefore not be amenable to the same systems for improvement for the hospital). Less than 24 hour admissions don't have the same risk for adverse events as children that are formally admitted to the pediatric unit. The nurses and other staff caring for them could be different as possibly are the reasons for the observation. This represents a reasonable exclusion especially when thinking about using this data for quality improvement and system improvement and determining the root cause of the events. Theoretically, SDS should not play a part in adverse events although patients and families can play a role in prevention of errors and health education level may be associated with better awareness of what is to be happening and lead to prevention (e.g. which medication is being given--the parent that asks the question may prevent the wrong one from being given). Surgical treatment is fundamentally different from medical treatment and therefore should be considered separately as planned. More details about the similarities and differences between the academic and community hospital might clarify whether or not that is an appropriate risk factor. Safety is an important part of quality care and measuring the number of preventable adverse events measures levels of safety and is a meaningful approach. Missing data/charts didn't seem to be an issue in the hospitals in which this was done, however, missing or unfinished charts might be traceable to a small group of practitioners and therefore might not be unbiased. From this study there didn't seem to be any bias in missing such a small number of charts to review.

Criterion 3. Feasibility

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

- The metric would require two primary reviewers, either nurses or physicians, to identify the trigger and adverse event. After identifying the trigger and adverse event, two physicians would then be required to "review each adverse event to determine whether that event may have been preventable." Each review is suggested to have a maximum time limit of a 30-minute review.
- The data elements would need to be abstracted from the record. The developer notes that "some data elements are defined in electronic sources."
- The developer notes that a completely automated process is not currently feasible, nor likely in the near future. As of now, automated trigger identification is possible, but has not been tested.

- The codes for the GAPPS automated triggers, the GAPPS Manual of Operations, and all associated forms that reviewers complete are available to users free of charge.
- The developer states that comparison of preventable AE rates across hospitals would require that reviewers receive adequate training in the trigger tool methodology to ensure standardization of the preventable AE detection process; the sensitivity analyses performed as part of validity testing reinforced this.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*
- *Is the data collection strategy ready to be put into operational use?*
- *While the education materials are of no cost, what are the costs associated with training a clinician to identify an event? Is there additional burden for that physician?*
- *Can trained nurses or nurse practitioners review records?*

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

RATIONALE: While the initial scan for triggers can be done by many types of clinicians, currently there is no automated format for the trigger tool. The assessment of the events preventability falls solely on physicians and requires two physicians with adequate training. This will result in increased burden on the physician population. The developer also notes that currently and in the near future, an automated process is not feasible.

Committee pre-evaluation comments

Criteria 3: Feasibility

- It was reassuring that the tools are available free. It is agreed that the data results are available however only if the medical records are accurate. It is agreed nurses could review and a barrier would be physician training. Rating: moderate.
- Some of the elements are likely available or searchable in an electronic record. Having a trigger list to look for particular events as well as a sampling strategy which does not pull a burdensome number of records to review, but with a pretty high pick up rate for adverse events is preferable to the other options which are reporting which significantly undercounts and chart review without the trigger list which is very resource intensive. This approach is used in the adult world successfully and there isn't any reason to expect that it can't be moved to the pediatric world. Training is required and that takes time for the physicians and other clinicians that might be doing the reviews. The primary reviewers could be nurses or nurse practitioners or physician assistants with the secondary reviewers being physicians. That would change some of the costs and nurses are commonly used to do chart reviews in a number of different environments.
- Concern that this measure requires manual review by at least two reviewers (physicians.) This would be a significant resource issue for hospitals, especially if they do not employ staff physicians who could do this work. However, due to the variation and complexity of inpatient pediatric cases, it might be appropriate to use physicians as reviewers to improve reliability.

Criterion 4: Usability and Use

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details

- This is a new measure, so it has not been used yet.

- The developer notes "AHRQ and CMS intend that the GAPPS measure be available for public use with the current expectation that the full measure specifications be provided on the AHRQ website, CMS website, or both." The developer also states "We anticipate that GAPPS results will be useful to everyone with a need for information on the quality of pediatric inpatient care, including patients, parents, hospitals, health plans, insurers, and policy makers. In addition, hospitals could provide GAPPS performance scores to quality organizations and purchasers."
- According to the developer, a "GAPPS Manual of Operations and trigger codes" and a series of training videos has been developed and are available online for sites that would like to utilize the measure.

Improvement results

- The developer expects using GAPPS to quantify inpatient adverse events will assist in identifying priorities and target available resources to help reduce preventable harm in children. The developer states that "Multiple studies have shown that hospitals with reliable means to track adverse events have experienced improvements in patient safety and associated clinical outcomes."

Unexpected findings (positive or negative) during implementation

- None reported by developer.

Potential harms

- None reported by developer.

Vetting of the measure

- This is a new measure and has not been implemented as of yet.

Feedback:

- This is a new measure and has not been implemented as of yet.

***Questions for the Committee**:*

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

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

Committee pre-evaluation comments

Criteria 4: Usability and Use

- It is concerning that that measure is neither publicly available, though intended to be so, nor in an accountability system. This measure will further the goal of high quality healthcare and not mentioned is cost savings. Rating: moderate. There is no section 5 but it is noted that this measure has not been vetted so is ineligible for endorsement and designation.
- Providing safer care with fewer errors not only is better for patients, but also saves the system money in the end. There are no downsides to this (except that it potentially uncovers adverse events and errors that weren't reported or that make a hospital system look worse than another one, but the goal is to use it to improve care). Looking at the lessons learned in the adult setting and how they approached this might be helpful to make this useable and feasible.
- While the measure would likely promote quality improvement efforts in hospitals, it would likely be difficult for purchasers/public to interpret any variation in results between hospitals.

Criterion 5: Related and Competing Measures

Related or competing measures

- 0715: Standardized adverse event ratio for children < 18 years of age undergoing cardiac catheterization (Under Annual Update)

Harmonization

- N/A

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

RATIONALE IF NOT ELIGIBLE: The measure has not yet been vetted by those being measured or other users.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: GAPPS: Rate of preventable adverse events per 1,000 patient-days among pediatric inpatients
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Not Applicable
Date of Submission: 12/7/2016

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

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

Notes

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

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome
☒ Health outcome: Click here to name the health outcome
☐Patient-reported outcome (PRO): Click here to name the PRO

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

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

☐ Process: Click here to name what is being measured

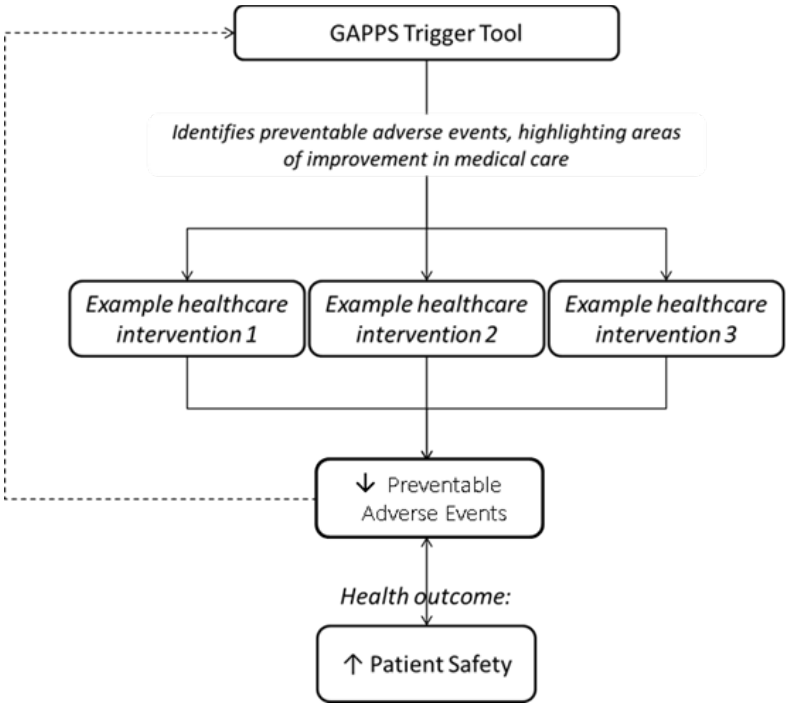
☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

GAPPS is a measure of global patient safety that specifically focuses on identifying preventable adverse events. All the AEs that GAPPS identifies represent instances in which existing patient safety interventions could be applied to prevent patient harm. GAPPS improves the capacity of hospitals to identify preventable AEs, target resources, and implement methods for decreasing risk of AEs in these prioritized domains. For example, GAPPS may assist hospitals in identifying preventable AEs resulting from errors. A hospital may implement new protocols to decrease the use of narcotics in an attempt to address preventable AEs related to the Naloxone (Narcan) administration trigger (see medication administration records [MARs] section of the included trigger list). Another example might be the initiation of rapid response teams to decrease AEs related to the trigger transfer to higher level of care (see physician orders section of trigger list). The advantage of the GAPPS Measure is that it allows hospitals to make notable improvements in patient safety while complementing voluntary event reporting. Using GAPPS, hospitals can implement evidence-based approaches to improve patient safety and measure the effectiveness of these efforts, allowing for ongoing improvement. Additionally, the GAPPS Measure can be used for analyses across institutions that may serve as the foundation for development of new evidence-based interventions to reduce preventable AEs.



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

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

Use of trigger tools to identify patient harm

The GAPPS measure uses a trigger tool to detect preventable AEs in hospitalized children. Multiple studies suggest that trigger tools are valid and reliable for tracking the incidence of patient harms in hospital settings, including within pediatric populations; trigger tools, including automated approaches to trigger tools, are better able to detect AEs than other methods (e.g., traditional voluntary incident reporting and detection tools used with administrative databases).(1–12) Trigger tool studies have identified specific triggers that have high yields for AE detection, such as "return to surgery," "positive blood culture," or "abrupt medication stop."(8,13,14) These findings were applied when developing the triggers used in the GAPPS measure.

In addition, nearly all studies evaluating mean medical record review times reported times under 30 minutes,(2,15–17) indicating that trigger tools can be used to adequately detect AEs with a reasonably small time burden. The GAPPS methodology likewise employs a maximum 30-minute time frame. Since GAPPS can be applied using either a manual or automated approach (the difference is whether triggers are initially identified by a primary reviewer or by an algorithm programmed into an electronic health record [EHR] system), the automated approach to identifying preventable AEs may decrease the time burden even further as shown in several studies.(18–21) As previously indicated, the automation of the trigger identification system has no impact on the measure beyond changing the means by which triggers are identified. The remaining pieces of the measure process following trigger identification are exactly the same for the automated and manual trigger approaches.

Relationship between patient safety and other aspects of quality

Some studies in adult populations have shown that patient harm rates are associated with other aspects of clinical quality. These studies demonstrate that rates of patient harm directly correlate with other quality metrics, including performance on clinical processes of care and other health outcomes.(22–26) For example, evidence shows that patients who experience healthcare-related harms have greater odds of in-hospital and 30-day mortality, as well as 30-day readmission.(23–26) It is likely that the same general associations hold true for pediatric populations.

Measuring patient safety to drive quality improvement

Patient safety is a core domain of healthcare quality and a major focus for quality improvement efforts.(10,27) Hospitals have been able to demonstrate that having a reliable means to track AEs leads to improvements in patient safety and associated clinical outcomes.(28–32) For example, hospitals that institute real-time adverse drug event surveillance systems are able to intervene before AEs become severe, or are able to prevent future AEs altogether.(28,30) In addition, the ability to track AEs allows for the design, implementation, and evaluation of targeted interventions, resulting in fewer AEs and decreased mortality.(29,30,32)

REFERENCES

1. Classen DC, Pestotnik SL, Evans RS, Burke JP. Computerized surveillance of adverse drug events in hospital patients. JAMA J Am Med Assoc. 1991 Nov 27;266(20):2847–51.

2. Agarwal S, Classen D, Larsen G, Tofil NM, Hayes LW, Sullivan JE, et al. Prevalence of adverse events in pediatric intensive care units in the United States. Pediatr Crit Care Med J Soc Crit Care Med World Fed Pediatr Intensive Crit Care Soc. 2010 Sep;11(5):568–78.

3. Call RJ, Burlison JD, Robertson JJ, Scott JR, Baker DK, Rossi MG, et al. Adverse Drug Event Detection in Pediatric Oncology and Hematology Patients: Using Medication Triggers to Identify Patient Harm in a Specialized Pediatric Patient Population. J Pediatr. 2014 Apr 24;

4. Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. “Global Trigger Tool” Shows That Adverse Events In Hospitals May Be Ten Times Greater Than Previously Measured. Health Aff (Millwood). 2011 Apr 1;30(4):581–9.

5. Griffin FA, Classen DC. Detection of adverse events in surgical patients using the Trigger Tool approach. Qual Saf Health Care. 2008 Aug 1;17(4):253–8.

6. Hooper AJ, Tibballs J. Comparison of a Trigger Tool and voluntary reporting to identify adverse events in a paediatric intensive care unit. Anaesth Intensive Care. 2014 Mar;42(2):199–206.

7. Kennerly DA, Kudyakov R, da Graca B, Saldaña M, Compton J, Nicewander D, et al. Characterization of Adverse Events Detected in a Large Health Care Delivery System Using an Enhanced Global Trigger Tool over a Five-Year Interval. Health Serv Res. 2014 Mar 13;

8. Rozich JD, Haraden CR, Resar RK. Adverse drug event trigger tool: a practical methodology for measuring medication related harm. Qual Saf Health Care. 2003 Jun 1;12(3):194–200.

9. Takata GS, Mason W, Taketomo C, Logsdon T, Sharek PJ. Development, testing, and findings of a pediatric-focused trigger tool to identify medication-related harm in US children’s hospitals. Pediatrics. 2008 Apr;121(4):e927-935.

10. Levinson DR. Adverse Events in Hospitals: Methods for Identifying Events. Washington, DC: Office of the Inspector General, Department of Health and Human Services; 2010. Report No.: OEI-06-08-00221.

11. Stockwell DC, Kirkendall E, Muething SE, Kloppenborg E, Vinodrao H, Jacobs BR. Automated adverse event detection collaborative: electronic adverse event identification, classification, and corrective actions across academic pediatric institutions. J Patient Saf. 2013 Dec;9(4):203–10.

12. Ferranti J, Horvath MM, Cozart H, Whitehurst J, Eckstrand J, Pietrobon R, et al. A Multifaceted Approach to Safety: The Synergistic Detection of Adverse Drug Events in Adult Inpatients. J Patient Saf. 2008 Sep;4(3):184–90.

13. Naessens JM, O’Byrne TJ, Johnson MG, Vansuch MB, McGlone CM, Huddleston JM. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. Int J Qual Health Care J Int Soc Qual Health Care ISQua. 2010 Aug;22(4):266–74.

14. Resar RK, Rozich JD, Simmonds T, Haraden CR. A trigger tool to identify adverse events in the intensive care unit. Jt Comm J Qual Patient Saf Jt Comm Resour. 2006 Oct;32(10):585–90.

15. Lander L, Roberson DW, Plummer KM, Forbes PW, Healy GB, Shah RK. A trigger tool fails to identify serious errors and adverse events in pediatric otolaryngology. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg. 2010 Oct;143(4):480–6.

16. Sharek PJ, Horbar JD, Mason W, Bisarya H, Thurm CW, Suresh G, et al. Adverse events in the neonatal intensive care unit: development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. Pediatrics. 2006 Oct;118(4):1332–40.

17. Unbeck M, Schildmeijer K, Henriksson P, Jürgensen U, Muren O, Nilsson L, et al. Is detection of adverse events affected by record review methodology? an evaluation of the “Harvard Medical Practice Study” method and the “Global Trigger Tool.” Patient Saf Surg. 2013;7(1):10.

18. Kaiser AM, de Jong E, Evelein-Brugman SF, Peppink JM, Vandenbroucke-Grauls CM, Girbes AR. Development of trigger-based semi-automated surveillance of ventilator-associated pneumonia and central line-associated bloodstream infections in a Dutch intensive care. Ann Intensive Care. 2014 Dec 21;4:40.

19. Jha AK, Kuperman GJ, Teich JM, Leape L, Shea B, Rittenberg E, et al. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. J Am Med Inform Assoc JAMIA. 1998 Jun;5(3):305–14.

20. Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripcsak G. Detecting Adverse Events Using Information Technology. J Am Med Inform Assoc. 2003 Mar 1;10(2):115–28.

21. Kirkendall ES, Spires WL, Mottes TA, Schaffzin JK, Barclay C, Goldstein SL. Development and Performance of Electronic Acute Kidney Injury Triggers to Identify Pediatric Patients at Risk for Nephrotoxic Medication-Associated Harm. Appl Clin Inform. 2014 Apr 2;5(2):313–33.

22. Calder L, Tierney S, Jiang Y, Gagné A, Gee A, Hobden E, et al. Patient safety analysis of the ED care of patients with heart failure and COPD exacerbations: a multicenter prospective cohort study. Am J Emerg Med. 2014 Jan;32(1):29–35.

23. Joice GA, Deibert CM, Kates M, Spencer BA, McKiernan JM. "Never events": Centers for Medicare and Medicaid Services complications after radical cystectomy. Urology. 2013 Mar;81(3):527–32.

24. Rosen AK, Geraci JM, Ash AS, McNiff KJ, Moskowitz MA. Postoperative adverse events of common surgical procedures in the Medicare population. Med Care. 1992 Sep;30(9):753–65.

25. Rosen AK, Loveland S, Shin M, Shwartz M, Hanchate A, Chen Q, et al. Examining the impact of the AHRQ Patient Safety Indicators (PSIs) on the Veterans Health Administration: the case of readmissions. Med Care. 2013 Jan;51(1):37–44.

26. Sukumar S, Roghmann F, Trinh VQ, Sammon JD, Gervais M-K, Tan H-J, et al. National trends in hospital-acquired preventable adverse events after major cancer surgery in the USA. BMJ Open. 2013;3(6).

27. Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human: Building a Safer Health System. Washington, DC: Institute of Medicine Committee on the Quality of Health Care in America; 1999.

28. Evans RS, Pestotnik SL, Classen DC, Bass SB, Burke JP. Prevention of adverse drug events through computerized surveillance. Proc Annu Symp Comput Appl Sic Med Care Symp Comput Appl Med Care. 1992;437–41.

29. Kaufman J, Rannie M, Kahn MG, Vitaska M, Wathen B, Peyton C, et al. An interdisciplinary initiative to reduce unplanned extubations in pediatric critical care units. Pediatrics. 2012 Jun;129(6):e1594-1600.

30. Piontek F, Kohli R, Conlon P, Ellis JJ, Jablonski J, Kini N. Effects of an adverse-drug-event alert system on cost and quality outcomes in community hospitals. Am J Health-Syst Pharm AJHP Off J Am Soc Health-Syst Pharm. 2010 Apr 15;67(8):613–20.

31. Cohen MM, Kimmel NL, Benage MK, Cox MJ, Sanders N, Spence D, et al. Medication safety program reduces adverse drug events in a community hospital. Qual Saf Health Care. 2005 Jun;14(3):169–74.

32. Garrett PR Jr, Sammer C, Nelson A, Paisley KA, Jones C, Shapiro E, et al. Developing and implementing a standardized process for global trigger tool application across a large health system. Jt Comm J Qual Patient Saf Jt Comm Resour. 2013 Jul;39(7):292–7.

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

Not applicable.

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Not applicable.

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

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

Not applicable.

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

1a.4.2 What process was used to identify the evidence?
Not applicable.

1a.4.3. Provide the citation(s) for the evidence.
Not applicable.

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

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

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

NQF_evidence_attachment_FINAL_20161206.docx

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

Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

No

1b. Performance Gap

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

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

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

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

IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.

Patient safety is a core domain of healthcare quality and a major focus for quality improvement efforts.(1,2) GAPPS is the first and only available global patient safety measure tailored for pediatric populations. By measuring preventable adverse event rates (i.e., harm) in inpatient pediatric populations, it provides important information to providers, hospital quality teams, and state health departments about outcomes of their patient care. Use of the measure will benefit patients, families, and providers because it enables stakeholders to identify and target areas of patient care that may benefit from quality improvement initiatives. Since GAPPS focuses on preventable adverse events, hospitals are able to assess and prioritize clinical areas with potential for immediate improvement.

The GAPPS Measure represents an opportunity to fill a notable void in safety measurement, specifically in pediatric preventable adverse events. National progress on quantifying and tracking AEs has stagnated because of the absence of an accepted national standard in all but a few defined areas. The GAPPS Measure can represent that national standard for pediatrics, and can be an important measure to evaluate and improve adverse event rates across different sites and time periods.

IMPORTANCE OF MEASURING ADVERSE EVENTS

Studies show that there is a high prevalence of medical errors and/or adverse events among the patient population. For example, in 1999, the Institute of Medicine (IOM) estimated that medical errors contribute up to 98,000 deaths and one million injuries each year.(2) In 2010, the Department of Health and Human Services' Office of the Inspector General estimated that 180,000 deaths due partly to adverse events occur among Medicare patients annually, making adverse events the third leading cause of death in the United States after heart disease and cancer.(3,4) Hospitalized pediatric patients, who tend to have unique diseases and care distinct from adult patients, are also vulnerable to high adverse event rates: published studies report 11.1 adverse drug events per 100 inpatient pediatric patients, 74 adverse events per 100 neonatal intensive care unit (NICU) patients, and 203 adverse events per 100 pediatric intensive care unit (PICU) patients.(5–9) Consequently, tracking adverse events in hospital settings is an important step towards understanding the current state of clinical care and creating initiatives aimed towards improving clinical quality.

Measuring preventable adverse event rates may also help hospitals better understand different aspects of their clinical quality. Some studies have found correlations between patient harm and other quality aspects such as performance on clinical processes of care and other health outcomes.(10–14) For instance, studies suggest that patients who experience healthcare-related harms

have greater odds of in-hospital and 30-day mortality, as well as 30-day readmission.(11–14) As such, measuring preventable adverse event rates is an essential first step for hospitals to understand and improve their patient care.

IDENTIFICATION OF INPATIENT ADVERSE EVENTS

Various approaches exist for identifying adverse events. Voluntary passive reporting systems are commonly employed but recognized to have low sensitivity.(15) A more reliable, sensitive methodology for capturing data on the safety of hospital care is thus essential.(16,17)

In 2003, AHRQ released its Patient Safety Indicators (PSI), developed in response to a congressional mandate to reduce medical errors.(18) PSIs are intended to identify events that most likely resulted from preventable medical errors.(18–20) PSIs have been used with some success but have a number of limitations, in part due to their reliance on administrative data. They have also been found to have low sensitivity.(17,18)

Use of trigger tools has been shown to be a faster, more sensitive, and more reliable method of adverse event detection than other approaches.(5,9,17,21–23) "Triggers" are red flags in a medical record that may indicate the presence of an underlying adverse event and prompt further inspection to determine whether an adverse event occurred.(21,22) An example trigger is the documented administration of an antidote-type medication (e.g., naloxone). Once a trigger is found, an in-depth review is undertaken to determine whether an adverse event occurred. In the case of naloxone, administration may indicate an adverse event occurred if the drug was given to counteract on overdose of opioids given in the hospital but may not if the overdose occurred due to voluntary recreational opioid use. Trigger tools detect adverse events in a high percentage of hospitalizations, ranging in published reports from 19% to 63%, and have evolved significantly over time.10-12

The Global Trigger Tool for Measuring Adverse Events (GTT), developed by the Institute for Healthcare Improvement (IHI), has become widely accepted as an effective approach for identifying adverse events in hospitalized adult patients.(2,9,15,23–26) The GTT approach identifies 10 times more adverse events than AHRQ's PSIs and almost 100 times more events than voluntary reporting.(1,17) However, the GTT has an exclusion of patients under age 18 so does not work for a pediatric population.

PEDIATRIC INPATIENT PATIENT SAFETY: LACK OF STANDARDIZED QUALITY MEASUREMENT

Although one study determined that a version of the GTT applied to the pediatric population could identify pediatric adverse events, the authors and other experts called for development of a standardized pediatric tool that focuses specifically on the problems of hospitalized children and that encompasses the breadth of inpatient pediatric care.(12, 18, 25, 29) The absence of a comprehensive pediatric trigger tool is a recognized limitation in quantifying the full scope of pediatric adverse events. An early effort to develop a pediatric-focused trigger tool led to the development of the Canadian Pediatric Trigger Tool.(30, 31)

We developed GAPPS to meet the need for a comprehensive, sensitive measure of pediatric patient safety. Our focus was on developing a global trigger tool for pediatric patients that could be more reliably applied across different hospital sites, both academic and community, than previous efforts. In addition, we sought to further refine the list of triggers to make a more robust global trigger list. We used methods similar to those used for GTT, including review of published tools and manual medical record review by experts in patient safety, which has been demonstrated to be a crucial component of developing patient safety measures.(30) We also utilized the RAND/UCLA Appropriateness Method, a 16-center field study, and post-analysis refinement of the trigger list to ensure GAPPS includes a more comprehensive trigger list than previous trigger tools.

GAPPS offers an enhancement in trigger tool methodology in that, unlike GTT, it requires that reviewers assess preventability. In the five years since IHI released the second edition of GTT, patient safety experts and national fiscal and quality improvement policies have increasingly focused on addressing preventable adverse events. The GAPPS measure uses the same approach to rate preventability as the North Carolina Patient Safety study, which was found to ascertain preventability with a high degree of reliability.(31)

DISPARITIES IN RISK OF HARM

Children with special healthcare needs experience elevated rates of medical errors.(32) Among hospitalized pediatric patients, those with chronic conditions are at significantly higher risk for medical errors than those without chronic conditions.(32)

What is known about racial/ethnic disparities in patient safety, particularly among children, is limited.(33) Black and Hispanic newborns are at higher risk of birth trauma.(18) In addition, extrapolations from associations between race/ethnicity and known risk factors for harm suggest that Black and Hispanic children are likely at greater risk of harm than White children. Because severity and complexity of illness increase the risk of errors, and Black and Hispanic children are at higher risk for more complex conditions, these children are at greater risk for adverse events.(33) Emergency room visits are also associated with increased risk of adverse events, and Black and Hispanic children are known to visit the emergency department more frequently.(28,33)

COSTS OF GAPPS UPTAKE

As is the case with all new quality measures, we recognize that there may be concerns about the uptake of GAPPS due to hospital resource constraints. AE monitoring, like patient experience surveying, is not possible using administrative data and therefore requires additional resources. As previously discussed, intensive measurement of preventable AEs is one of the most important areas of quality measurement, one that has to this point been inadequately addressed in pediatric patient populations. The capacity of the GAPPS Measure to spur nationwide improvement in pediatric patient safety represents a significant potential return on investment, one that outweighs implementation resource concerns.

It is important to note that some of the resources and infrastructure to successfully implement the GAPPS measure are already in place, as states commit to systems of mandatory reporting of certain adverse events. About half of states currently have mandatory reporting systems in place, with many others reporting AEs on a voluntary basis.(34) GAPPS will require an augmentation of processes hospitals are already initiating to report AEs, but will yield far greater returns. Far from representing a duplicative burden, the GAPPS Measure can take advantage of recent prioritization of patient safety and the resources hospitals are putting in place to evaluate quality in this domain. The marginal resources used for GAPPS implementation provide a drastically more robust assessment of hospital safety than currently in place, as global trigger tools have been shown to capture up to ten times more AEs than alternative AE measurement methods.(17,35)

REFERENCES

1. Levinson DR. Adverse Events in Hospitals: Methods for Identifying Events. Washington, DC: Office of the Inspector General, Department of Health and Human Services; 2010. Report No.: OEI-06-08-00221.
2. Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human: Building a Safer Health System. Washington, DC: Institute of Medicine Committee on the Quality of Health Care in America; 1999.
3. James JT. A new, evidence-based estimate of patient harms associated with hospital care. J Patient Saf. 2013 Sep;9(3):122–8.
4. Allen M, ProPublica. How Many Die From Medical Mistakes In U.S. Hospitals? [Internet]. NPR.org. [cited 2014 Jun 14]. Available from: http://www.npr.org/blogs/health/2013/09/20/224507654/how-many-die-from-medical-mistakes-in-u-s-hospitals
5. Sharek PJ, Classen D. The incidence of adverse events and medical error in pediatrics. Pediatr Clin North Am. 2006 Dec;53(6):1067–77.
6. Pediatric Patient Safety Measurement Tool (PPSMT) Pilot Toolkit. Children's Hospital Association; 2012.
7. Stockwell DC, Slonim AD. Quality and safety in the intensive care unit. J Intensive Care Med. 2006 Aug;21(4):199–210.
8. Bates DW, Spell N, Cullen DJ, Burdick E, Laird N, Petersen LA, et al. The costs of adverse drug events in hospitalized patients. Adverse Drug Events Prevention Study Group. JAMA J Am Med Assoc. 1997 Jan 22;277(4):307–11.
9. Takata GS, Mason W, Taketomo C, Logsdon T, Sharek PJ. Development, testing, and findings of a pediatric-focused trigger tool to identify medication-related harm in US children's hospitals. Pediatrics. 2008 Apr;121(4):e927-935.
10. Calder L, Tierney S, Jiang Y, Gagné A, Gee A, Hobden E, et al. Patient safety analysis of the ED care of patients with heart failure and COPD exacerbations: a multicenter prospective cohort study. Am J Emerg Med. 2014 Jan;32(1):29–35.
11. Joice GA, Deibert CM, Kates M, Spencer BA, McKiernan JM. "Never events": Centers for Medicare and Medicaid Services complications after radical cystectomy. Urology. 2013 Mar;81(3):527–32.
12. Rosen AK, Geraci JM, Ash AS, McNiff KJ, Moskowitz MA. Postoperative adverse events of common surgical procedures in the Medicare population. Med Care. 1992 Sep;30(9):753–65.
13. Rosen AK, Loveland S, Shin M, Shwartz M, Hanchate A, Chen Q, et al. Examining the impact of the AHRQ Patient Safety Indicators (PSIs) on the Veterans Health Administration: the case of readmissions. Med Care. 2013 Jan;51(1):37–44.
14. Sukumar S, Roghmann F, Trinh VQ, Sammon JD, Gervais M-K, Tan H-J, et al. National trends in hospital-acquired preventable adverse events after major cancer surgery in the USA. BMJ Open. 2013;3(6).
15. Sharek PJ, Horbar JD, Mason W, Bisarya H, Thurm CW, Suresh G, et al. Adverse events in the neonatal intensive care unit: development, testing, and findings of an NICU-focused trigger tool to identify harm in North American NICUs. Pediatrics. 2006 Oct;118(4):1332–40.
16. Landrigan CP, Parry GJ, Bones CB, Hackbarth AD, Goldmann DA, Sharek PJ. Temporal trends in rates of patient harm resulting from medical care. N Engl J Med. 2010 Nov 25;363(22):2124–34.
17. Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. "Global Trigger Tool" Shows That Adverse Events In Hospitals May Be Ten Times Greater Than Previously Measured. Health Aff (Millwood). 2011 Apr 1;30(4):581–9.
18. Miller MR, Elixhauser A, Zhan C. Patient Safety Events During Pediatric Hospitalizations. Pediatrics. 2003 Jun 1;111(6):1358–66.
19. Dynan L, Goudie A, Smith RB, Fairbrother G, Simpson LA. Differences in Quality of Care Among Non–Safety-Net, Safety-Net, and Children's Hospitals. Pediatrics. 2013 Feb 1;131(2):304–11.
20. Miller MR, Zhan C. Pediatric Patient Safety in Hospitals: A National Picture in 2000. Pediatrics. 2004 Jun 1;113(6):1741–6.

21. Rozich JD, Haraden CR, Resar RK. Adverse drug event trigger tool: a practical methodology for measuring medication related harm. Qual Saf Health Care. 2003 Jun 1;12(3):194–200.
22. Resar RK, Rozich JD, Classen D. Methodology and rationale for the measurement of harm with trigger tools. Qual Saf Health Care. 2003 Dec 1;12(suppl 2):ii39-ii45.
23. Agarwal S, Classen D, Larsen G, Tofil NM, Hayes LW, Sullivan JE, et al. Prevalence of adverse events in pediatric intensive care units in the United States. Pediatr Crit Care Med J Soc Crit Care Med World Fed Pediatr Intensive Crit Care Soc. 2010 Sep;11(5):568–78.
24. Thomas EJ, Studdert DM, Burstin HR, Orav EJ, Zeena T, Williams EJ, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. Med Care. 2000 Mar;38(3):261–71.
25. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. CMAJ Can Med Assoc J J Assoc Medicale Can. 2004 May 25;170(11):1678–86.
26. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The Quality in Australian Health Care Study. Med J Aust. 1995 Nov 6;163(9):458–71.
27. Kirkendall ES, Kloppenborg E, Papp J, White D, Frese C, Hacker D, et al. Measuring adverse events and levels of harm in pediatric inpatients with the Global Trigger Tool. Pediatrics. 2012 Nov;130(5):e1206-1214.
28. Matlow AG, Baker GR, Flintoft V, Cochrane D, Coffey M, Cohen E, et al. Adverse events among children in Canadian hospitals: the Canadian Paediatric Adverse Events Study. CMAJ Can Med Assoc J J Assoc Medicale Can. 2012 Sep 18;184(13):E709-718.
29. Stockwell D, Bisarya H, Classen D, Kirkendall E, Landrigan C, Lemon V, et al. A trigger tool to detect harm in pediatric inpatient settings. Pediatrics. 2015;
30. Scanlon MC, Harris JM, Levy F, Sedman A. Evaluation of the Agency for Healthcare Research and Quality Pediatric Quality Indicators. Pediatrics. 2008 Jun 1;121(6):e1723–31.
31. Sharek PJ, Parry G, Goldmann D, Bones K, Hackbarth A, Resar R, et al. Performance characteristics of a methodology to quantify adverse events over time in hospitalized patients. Health Serv Res. 2011 Apr;46(2):654–78.
32. Ahuja N, Zhao W, Xiang H. Medical Errors in US Pediatric Inpatients With Chronic Conditions. Pediatrics. 2012 Oct 1;130(4):e786–93.
33. Flores G, Ngui E. Racial/Ethnic Disparities and Patient Safety. Pediatr Clin North Am. 2006 Dec;53(6):1197–215.
34. QuPS.org - Medical Errors and Patient Safety - US States Comparison - Overview - [Internet]. [cited 2016 Nov 16]. Available from: http://www.qups.org/med_errors.php?c=internal&id=195
35. Office of the Inspector General. Adverse Events In Hospitals: National Incidence Among Medicare Beneficiaries. Department of Health and Human Services,; 2010.

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

We conducted a National Field Test at 16 hospitals from across the United States that represented diverse geographic regions (four hospitals in each of the four US census regions: Northeast, South, West, and Midwest) and included eight teaching and eight non-teaching hospitals (teaching status was based on categorization set by the American Hospital Association). 3,790 hospitalizations occurring between 2007 and 2012 were included in this analysis.

Current rate of preventable AEs. Of the 414 AEs identified, 210 (50.7%) AEs were preventable, representing 9.5 preventable AEs [CI 8.2-10.8]/1,000 patient days. Compared to community hospitals, academic hospitals had higher preventable harm rates (13.1 [CI 11.4-15.2] vs. 2.4 [CI 1.5-3.8] AEs/1,000 patient days, $p<0.001$). GAPPS is a measure of preventable adverse events. Because the identified adverse events are preventable, the ideal would be to have no preventable AEs. The GAPPS Measure specifically focuses on preventable adverse events because it outlines areas for immediate improvement. As we will discuss in more detail later in the application, the GAPPS Measure indicates there is significant room for improvement across most institutions and patient demographics. In addition to widespread incidence of preventable AEs, there is dramatic variation across institutions (preventable AEs/1000 patient days ranged from 0-20.4) and significant subpopulation disparities for race, medical complexity, and insurance type. In sum, GAPPS demonstrates areas to target to reduce disparities for subpopulations that may be at higher risk of encountering a preventable AE.

Changes in the rate of preventable AEs over time. Multivariate analyses controlling for demographic characteristics and chronic conditions showed no significant changes in preventable AE rates over time. Poisson regression accounting for hospital-level clustering and changes over time found no significant changes over time in preventable AEs (risk factor=1.00/1,000 patient days [CI 0.98-1.02]. When stratified by hospital type, neither academic nor community hospitals experienced significant temporal trends in preventable AEs/1,000 patient days.

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

Not Applicable

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

RACE/ETHNICITY

We assessed differences in pediatric patient safety associated with race/ethnicity by evaluating whether the rate of preventable adverse events (AEs) identified by reviewers varied among racial/ethnic groups. Race/ethnicity was recorded in our National Field Test using the categories Alaska Native, American Indian, Asian, Black, Hispanic, Native Hawaiian or other Pacific Islander, Other, and White. For our analysis, we combined Alaska Native, American Indian, Asian, Native Hawaiian or other Pacific Islander, and non-White, non-Hispanic Other patients into a single "Other" category because each of the categories was represented by a very small number of hospitalizations.

Table 1 –Preventable AEs per 1,000 patient days by race/ethnicity (n = 3,231)

Race/Ethnicity	n (%)	prev AE rate	P-Value
White	2,152 (56.8%)	8.9	Reference
Black	476 (12.6%)	5.9	0.13
Hispanic	419 (11.1%)	15.9	0.002
Other	184 (4.9%)	11.1	0.47

Across all sites evaluated, we found that Hispanic patients had a higher unadjusted preventable AE rate at 15.9, compared to White patients at 8.9 (p=0.002).

PATIENTS WITH CHRONIC CONDITIONS

We assessed differences in pediatric patient safety associated with presence of chronic conditions by evaluating whether the rate of preventable AEs identified by reviewers varied based on the chronic conditions present among patients (as classified in AHRQ's CCI system), controlling for length of hospitalization.(38)

Table 2 – Preventable AEs per 1,000 patient days by numbers of Chronic Condition Indicators per patient (n= 3,524)

Chronic Condition Indicators	n (%)	prev AE rate	P-Value
0 body system	1,990 (52.5%)	6.5	Reference
1 body system	1,085 (28.6%)	9.5	0.04
2 body systems	321 (8.5%)	17.9	<0.001
3 or more body systems	128 (3.4%)	19.8	<0.001

Overall, we found that patients with a body system affected by a chronic condition had higher unadjusted preventable AE rates than those without any body system affected by a chronic condition. Particularly, patients with 3 or more body systems affected by a chronic condition had the highest unadjusted preventable AE rate at 19.8 (p<0.001) as shown in Table 2.

INSURANCE STATUS

We assessed differences in pediatric patient safety associated with socioeconomic status (SES) by using insurance status as a proxy for SES and examining whether the rate of preventable AEs identified by reviewers varied with insurance status. Insurance status was captured in our National Field Test using six non-mutually exclusive categories: Medicaid, Medicare, Private Insurance, Self-Pay, No Insurance, and Not Recorded. These sorted the cohort into eight unique categories (some of which indicate that a patient had multiple insurance types listed during the hospitalization included in our field test): no insurance; private insurance; public insurance; private insurance and self-pay; public and private insurance; public and no insurance; public, self-pay and private insurance; and insurance not recorded.

We chose to exclude hospitalizations for patients covered by Medicare from the analysis because pediatric eligibility for Medicare is based on having specific medical conditions rather than being based solely on family income.(39) Therefore, we evaluated patients with public insurance (Medicaid), private insurance, and no insurance. Patients who were recorded to have both private insurance and public insurance were categorized as patients with private insurance.

Table 3 shows the distribution of the insurance types included in our analysis. 37.5% of the patients had public insurance, 59.5% had private insurance, and 3.0% did not have insurance. 8.5% were missing insurance information and not included in this analysis. Patients with private insurance (preventable AE rate 8.5) and no insurance (preventable AE rate 3.9) had lower crude preventable AE rates across sites than patients with public insurance (AE rate 12.1).

Table 3 – Preventable AEs per 1,000 patient days by insurance type (n = 3,468)

Insurance	n (%)	prev AE rate	P-Value
Public	1,300 (37.5%)	12.1	Reference
Private	2,064 (59.5%)	8.5	0.02
No Insurance	104 (3.0%)	3.9	0.11

REFERENCES

38. Chronic Condition Indicator (CCI) for ICD-10-CM [Internet]. Healthcare Cost and Utilization Project (HCUP); Available from: https://www.hcup-us.ahrq.gov/toolssoftware/chronic_icd10/chronic_icd10.jsp
39. Medicare C for, Baltimore MS 7500 SB, Usa M. Medicare [Internet]. 2014 [cited 2014 Sep 2]. Available from: http://www.cms.gov/Medicare/Medicare.html

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

Not Applicable

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

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

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.childrenshospital.org/research-and-innovation/research/centers/center-of-excellence-for-pediatric-quality-measurement-cepqm/cepqm-measures/global-tool-of-patient-safety/content

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

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

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.
No

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.
Not Applicable

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The number of preventable adverse events found in a patient sample.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Adverse events are defined as "unintended physical injuries resulting from or contributed to by medical care that require additional monitoring, treatments, or hospitalizations, or that result in death."(1,2) This matches the Institute for Healthcare Improvement's adult Global Trigger Tool's (IHI GTT's) definition of harm since "harm" and "adverse event" are used synonymously in the context of patient safety.(1) GAPPS includes assessments of preventability to facilitate the identification of clinical areas with potential for immediate improvement.

The GAPPS measure requires two physicians to review and independently rate the preventability of each adverse event case they review. When physicians disagree on an event's preventability, they discuss the rationale for their ratings with one another until both agree on whether an adverse event is preventable or not. A third physician is consulted in the rare occasion that the two physicians continue to disagree on an event's preventability after discussing with one another.

REFERENCES
1. Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events (Second Edition). Institute for Healthcare Improvement; 2009. (IHI Innovation Series white paper).
2. Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, et al. "Global Trigger Tool" Shows That Adverse Events In Hospitals May Be Ten Times Greater Than Previously Measured. Health Aff (Millwood). 2011 Apr 1;30(4):581–9.

Below is a list of example triggers from the GAPPS Measure that are often found by reviewers in various sections of the medical record. For a full list of GAPPS triggers and a description of each, see appendix A.1.

Discharge summary
- All inpatient deaths
- Mechanical ventilation >48 hours
- Hospital readmission within 30 days
- Return to surgery

Laboratory reports
- Valproic acid >170 mcg/ml
- Carbamazepine >20 mcg/ml
- Serum creatinine doubling

- Nephrotoxin use (e.g., aminoglycosides, cyclosporine, tacrolimus, vancomycin) and rising creatinine (Cr)
- Hepatotoxic medications and elevated liver enzymes (AST, ALT)
- Drop of hemoglobin (Hgb) or hematocrit (Hct) of >25% in less than 24 hours

Radiology results
- Patient fall

Physician orders
- Abrupt medication stop
- Transfer to higher level of care

Medication administration records (MARs)
- Vitamin K administration after warfarin
- Naloxone administration
- Hypoglycemia (<2 mmol/L or 40 mg/dL)

Nursing flow sheets
- Surgical site infection
- Infiltration/phlebitis documentation
- Embolus/thrombus documentation
- Pressure ulcer documentation (= stage 2)

Procedure notes (diagnostic, surgical)
- Any code or arrest, or rapid response team activation
- Mechanical ventilation greater than 48 hours post-operative

Nursing/Physician/Multi-disciplinary progress notes
- Opiate-related constipation with intermittent laxative use
- Healthcare-associated infections: positive C. difficile test
- Healthcare-associated infections: positive blood culture (only after 48 hours from admission)
- Healthcare-associated infections: positive urine culture (only after 48 hours from admission)
- Healthcare-associated infections: positive respiratory or GI viral test (only after 48 hours from admission)
- Racemic epinephrine administration (patients mechanically ventilated within the last 24 hours)

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator is 1,000 patient-days for all sampled pediatric patients who meet inclusion, but not exclusion, criteria.

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The denominator includes all patients who meet the following criteria:
1. Patients <18 years of age at admission;
2. Patients with length of stay (LOS) greater than or equal to 24 hours;
3 Patients admitted for acute care. Acute care does not include patients discharged from the Emergency Department without admission to the hospital; or patients in rehabilitation and residential units, non-acute inpatient psychiatric units, newborn nurseries, and day treatment areas. If a patient is initially admitted acutely but subsequently transferred to inpatient psychiatric care, the acute portion of the hospitalization should be included; and
4. Patients who were discharged from, who were transferred out of, or who died during the inpatient or observation hospital stay.

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

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

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*
Stratification is not required within institutions. However, if desired, quality improvement teams may choose to stratify preventable adverse event rates. Variables commonly used to stratify outcome measures include service (e.g., medical versus surgical), department (e.g., cardiology, neurology, etc.), and patient safety focus area (e.g., healthcare-associated infections).

For comparisons between institutions, preventable adverse event rates should be stratified by teaching versus community hospitals due to differences in types (e.g., complexity) of patient populations

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
Statistical risk model
If other:

S.12. **Type of score:**
Rate/proportion
If other:

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

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
GAPPS allows quality improvement teams to measure preventable adverse event rates over time among pediatric inpatients. GAPPS can be applied within entire hospitals, individual divisions or services, or specific programs. For more detailed instructions on how to find preventable adverse events using either GAPPS' manual or automated approach, refer to Appendix A.

Step 1 – Assemble a review team
The GAPPS review team should consist of:

- Two primary reviewers who are responsible for reviewing and identifying adverse events in medical records. The second primary reviewer will only review a subset of the first primary reviewer's charts for a reliability check. It is recommended that each primary reviewer have extensive clinical experience, have familiarity with multiple clinical settings and interventions (including diagnostic tests, medications, and procedures), and be well-acquainted with the hospital's medical record system and typical delivery of care. The primary reviewer in trigger tool applications has historically been a nurse, but physicians, physician assistants, and pharmacists – among others –may also be good candidates.
- Two secondary reviewers who are responsible for reviewing any suspected adverse event identified by a primary reviewer. The secondary reviewers verify the occurrence of adverse events, as well as the ratings of severity and preventability for the events. They do not review medical records directly; instead, they listen to the primary reviewer's description of the adverse events he or she identified and ask questions as needed for clarification. Some secondary reviewers may choose to read the primary reviewer's written assessment in addition to listening to the reviewer's description of the hospitalization. Secondary reviewers should be physicians.

Step 2 – Select relevant hospitalizations
We recommend that the main primary reviewer selects a random sample of at least 20 inpatient hospitalizations each month from a list of all inpatient hospitalizations with discharge dates that fall within the month being reviewed; the hospitalizations may be drawn from an entire hospital or from a specific division, service, or program. The hospitalizations should meet eligibility criteria (noted below) for a minimum of 60 hospitalizations per quarter. For institutions with high pediatric patient volume, records for 60 unique patients typically will be reviewed. However, patients who have multiple discharges that fall within a given quarter may have their records reviewed multiple times.

A two-stage process is used to determine which pediatric medical records should be included in the GAPPS sample frame. The first stage determines whether patients meet the inclusion criteria listed below. For patients who meet inclusion criteria, certain exclusion criteria – also described below, are then applied.

Inclusion Criteria:
GAPPS is intended for broadly reviewing the medical records of pediatric patients who meet the following criteria:

- Patients <18 years of age at admission;
- Patients with length of stay (LOS) =24 hours;
- Patients admitted for acute care. Acute care does not include patients discharged from the Emergency Department without admission to the hospital; or patients in rehabilitation and residential units, non-acute inpatient psychiatric units, newborn nurseries, and day treatment areas. If a patient is initially admitted acutely but subsequently transferred to inpatient psychiatric care, the acute portion of the hospitalization should be included; and
- Patients who were discharged from, who were transferred out of, or who died during the inpatient or observation hospital stay.

Exclusion Criteria:
Patients with inpatient LOS <24 hours are excluded because patients with brief hospital stays are less likely to have received the amount of medical intervention necessary to evaluate the quality of care.

Patients =18 years of age at admission are excluded because the Center of Excellence for Pediatric Quality Measurement's (CEPQM) task was to create a tool for measuring patient safety in the pediatric age group (i.e., <18 years of age). With this in mind, GAPPS is designed to perform exclusively in pediatric patients.

Step 3 – Review of patient records by primary reviewers and secondary reviewers
Primary reviewers should spend up to 30 minutes reviewing each hospitalization in a medical record. They should focus on identifying and recording triggers and adverse events (for lists of the GAPPS manual and automated triggers, see Appendix A).

- Identifying triggers: When a trigger is discovered in the record (either manually or automatically via an electronic health record (EHR) system that flags hospitalizations), primary reviewers should look for information relevant to that trigger to investigate whether an adverse event occurred. Reviewers typically identify many more triggers than adverse events. If no adverse event is found, continue reviewing the remainder of the record for additional triggers. The manner in which the trigger is identified (manually or automatically) has no impact on the rest of the GAPPS measure process. The automated trigger list removes the arduous human identification factor from the process, but the measure remains exactly the same following trigger identification.
  - Some adverse events will be found without the identification of a related trigger. These events should still be recorded in the Primary Review Forms and Suspected Adverse Event Forms.
- Identifying adverse events: Whether discovered due to a positive trigger or encountered while searching for triggers, adverse events and their corresponding information should be recorded by the primary reviewer. We recommend that reviewers consider the following items when determining whether an adverse event has occurred:
  - Harm likely occurred through event(s) in which people experiencing the event would be unhappy the event occurred (e.g., IV infiltrate, even if minor).
  - Adverse events are, by definition, the result of medical treatment. If an incident was part of the natural progression of a patient´s disease process, it is unlikely to be an adverse event (e.g., patient admitted for respiratory failure due to pneumonia worsens despite appropriate management and consequently needs to be intubated), unless medical care somehow contributed to the incident.
  - Incidents that are the intended results of medical care are not considered adverse events (e.g., neutropenia with chemotherapy).
  - Psychological harm alone is not generally considered an adverse event (e.g., stress).

All identified adverse events should be recorded, regardless of location. The Primary Review Forms and Suspected Adverse Event Forms allow reviewers to specify where harms occurred, so harms occurring outside the hospital can be analyzed separately or removed from assessments of unit/hospital care quality as needed.

- Determining severity
  - Severity: Reviewers should assign severity to an adverse event using the five-point severity scale below, which is a modified version of the National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) Index for Categorizing Errors. Since the categories are not mutually exclusive, reviewers should assign the highest severity category that applies to the adverse event. It is important to note that adverse events in high-severity categories do not have to meet all of the requirements of lower-harm-level categories. For example, an adverse event can be categorized in harm level H (i.e., insulin bolus) but not qualify as a G-level harm (i.e., permanent injury).

Category E: Temporary harm to the patient and required intervention
Category F: Temporary harm to the patient and required initial or prolonged hospitalization
Category G: Permanent patient harm
Category H: Intervention required to sustain life
Category I: Patient death

Step 4- Determine preventability of adverse event
Primary reviewers (nurses) record preventability for data collection and internal validity assessment purposes. However, the final determination of preventability is made by the secondary reviewers (physicians). All reviewers should rely on the category definitions provided below and their own clinical experience when determining preventability. Training sessions, discussions with the review team, and experience with reviews will be crucial in developing consistent preventability ratings.

Categories of Preventability [1]
- Definitely not preventable: Events in which no obvious error occurred; necessary precautions were taken; no alteration in method or care exists to prevent the event.
  - Drug-associated rash (no prior exposure or history): A 9-year-old male with no known allergies presented to the emergency department for a sore throat, cough, and fever. When the patient was given ibuprofen for his fever, he developed hives and itching. The patient was then given diphenhydramine and responded well to the drug with no respiratory distress. Ibuprofen was discontinued and listed as an allergy on the patient's medical record.
- Probably not preventable: Events that do not appear preventable but would require further investigation to assess certainty.
  - Procedural complications (with skilled proceduralist and no errors): Despite nursing standards being followed, a 7-year-old female developed an IV infiltrate.
- Probably preventable: Events that appear preventable but would require further investigation to assess certainty.
  - Hospital-acquired infections: A male infant born at 35 weeks estimated gestation age had an umbilical catheter placed. An inflamed wound developed at the catheter site, and he was started on antibiotics. An abscess formed at the site over the next few days, so the wound was drained, and cultures were obtained that were positive for MRSA and Enterobacter spp.
- Definitely preventable: Events where error was identified; necessary precautions were not taken; event was preventable by modification of behavior, technique, or care.
  - Medication overdose: A 13-year-old female was given an overdose of insulin during treatment for diabetic ketoacidosis. Her blood glucose dropped precipitously, and she required a D50 bolus.

[1] While secondary reviewers can select one of four preventability rankings for each adverse event, preventability rankings are categorized into two groups when assessing secondary reviewer agreement and during data analysis. Specifically, adverse events ranked as "definitely not preventable" and "probably not preventable" are considered "nonpreventable," and adverse events ranked as "definitely preventable" and "probably preventable" are considered "preventable."

Step 5 – Record data in appropriate forms
Primary reviewers
Primary reviewers should complete the Primary Review Form for each hospitalization. For each adverse event, they should also complete the Suspected Adverse Event Form.

Secondary reviewers
Secondary Reviewer A should complete the Secondary Review Form A for each suspected adverse event identified by a primary reviewer, either confirming or denying that an adverse event occurred. Secondary Reviewer B should complete the Secondary Review Form B for each suspected adverse event identified by the primary reviewers, either confirming or denying that an adverse event occurred.

In cases in which Secondary Reviewers A and B disagree about whether an adverse event occurred or do not independently rate an adverse event with the same severity and preventability (note: preventability agreement is determined dichotomously, i.e., definitely/probably preventable vs. definitely/probably not preventable), the secondary reviewers must discuss the issues and reach consensus on all rankings. If the two secondary reviewers are unable to reach a consensus after discussing the case, a third physician should be consulted. Once reviewers agree on all rankings, one of the reviewers should complete the Consensus Form.

Step 6 – Check reliability

To assess the reliability with which institutions use GAPPS to identify triggers and adverse events, a second primary reviewer should perform a completely independent review of a random 10% sample of the medical records reviewed by the main primary reviewer from each sampling time frame (i.e., 6 records per quarter). This second review should occur at the end of each year on a total of 24 records annually. During this check, the second primary reviewer completes the same forms as the first primary reviewer: the Primary Review Form and, for each adverse event identified in a medical record, the Suspected Adverse Event Form. Knowing the rates at which primary reviewers identify and agree about adverse events will allow institutions to assess the reliability of their adverse event detection and to improve training efforts for reviewers as needed.

Step 7 – Analyze data
After the primary and secondary reviewers complete their reviews in each collection period, the data should be analyzed by computing preventable adverse events per 1,000 patient-days using the following equation: [(Total number of preventable adverse events identified in all the medical records in the sampling frame)/(Sum of the total number of inpatient days for all of the medical records reviewed in the sampling period)]*1,000. When comparing across institutions, the unit of time should be annual.

Case-mix adjustment for inter-hospital comparisons:
We recommend groups use mixed effects negative binomial regression to adjust preventable adverse event rates based on patient characteristics and type of service. Specifically, the outcome is the number of preventable adverse events for an admission (exposure time equal to length of stay), case-mix variables are fixed effects, and a hospital-level random intercept represents the variation between hospitals. Case-mix models should be stratified by hospital type (teaching vs. community). The case-mix data are obtained from the Primary Review Forms.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Hospitals and departments using GAPPS to measure preventable adverse event rates are responsible for generating complete, accurate, and valid lists of all pediatric inpatient hospitalizations (<18 years old) discharged between the first and last days of each month (e.g., for January, any qualifying discharges between and including the 1st and 31st days). The hospitalizations may be drawn from an entire hospital or from a specific division, service, or program.

We recommend that primary reviewers select a random sample of at least 20 inpatient hospitalizations each month from a list of all inpatient hospitalizations with discharge dates that fall within the month being reviewed. The hospitalizations should meet eligibility criteria (see Appendix A) for a minimum of 60 hospitalizations per quarter. We recommend a minimum sample size of 60 records per quarter in order for institutions to achieve adequate reliability for estimates of hospital-level preventable adverse event incidence. This sample size is based on the assumption that the trigger tool will be used in an improvement setting, for which the aim is to detect trends in the data showing meaningful change over time. According to Perla and colleagues, to plot the data quarterly, the appropriate sample size of medical records is given by 9/R, where R is the average number of adverse events per person.(3) Assuming an adverse event rate of at least 0.15, the recommended sample size computes to 9/0.15=60.

For institutions with high pediatric patient volume, records for 60 unique patients will typically be reviewed. However, patients who have multiple discharges that fall within a given quarter may have their records reviewed multiple times.

Records should be selected through a random process to eliminate any potential bias. A variety of selection methods can be used to ensure a random sample, such as:

- For a given month, number all discharge records (including those ending with deaths) sequentially starting with one. Using random-number generating software, generate 25 numbers between one and the total number of records. Select the hospitalizations labeled with the random number. Keep the first 20 hospitalizations that meet eligibility criteria. If 25 records are insufficient to yield 20 eligible hospitalizations, select more hospitalizations using the random-number generating software until you obtain the 20 that are needed.

For medical facilities, departments, or programs that have fewer than 20 pediatric inpatients per month, review records for all hospitalizations during the month.

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Not Applicable

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Electronic Health Record (Only), Paper Records

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Primary Review Form, Suspected Adverse Event Form, Secondary Review Form A Secondary Review Form B, Consensus Form

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

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

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Hospital : Acute Care Facility
If other:

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

2. Validity – See attached Measure Testing Submission Form

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.3 For maintenance of endorsement
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

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

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

What were the statistical results of the analyses used to select risk factors?

Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of

between-unit effects and within-unit effects)

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: GAPPS: Rate of preventable adverse events per 1,000 patient-days among pediatric inpatients

Date of Submission: 12/7/2016

Type of Measure:

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

Instructions

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

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

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

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

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

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

the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

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

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

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

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

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

Notes

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

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

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

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

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

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

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

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

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

Measure Specified to Use Data From:	Measure Tested with Data From:

(*must be consistent with data sources entered in S.23*)	
☒ abstracted from paper record	☒ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☒ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
Not applicable

1.3. What are the dates of the data used in testing? 2007-2012
The hospitalizations used in testing GAPPS have discharge dates that fall between 2007 and 2012.

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

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)
During the National Field Test, GAPPS was used at 16 hospital study sites. The hospitals were situated across the United States, representing diverse geographic regions (four hospitals in each of the four US census regions: Northeast, South, West, Midwest) and included eight teaching and eight non-teaching hospitals (teaching status was based on standards set by the American Hospital Association). Hospitals were identified to participate through the Pediatric Research in Inpatient Settings (PRIS) network. During the range of years in which we were sampling patient records (2007-2012) each hospital site was at a different stage in the evolution from paper to electronic records, with varying use of electronic health records.

Table 1 - Hospital Characteristics:

Hospital Characteristics	**Hospitals (Total N = 16)**
Teaching Status	
Teaching	8
Non-Teaching	8
Regions	
Northeast	4
Midwest	4
South	4

West	4
Type	
Free-Standing	5
General Population	11

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)?
(*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)
We reviewed 3,814 medical records for pediatric patients with discharge dates from January 1, 2007 to December 31, 2012 across the 16 hospital sites (≈240 records/hospital). Sampled patients who met the following criteria were included in the study:

- Patients <18 years of age at admission;
- Patients with length of stay (LOS) =24 hours;
- Patients admitted for acute care. Acute care does not include patients discharged from the Emergency Department without admission to the hospital; or patients in rehabilitation and residential units, non-acute inpatient psychiatric units, newborn nurseries, and day treatment areas. If a patient is initially admitted acutely but subsequently transferred to inpatient psychiatric care, the acute portion of the hospitalization should be included; and
- Patients who were discharged from, who were transferred out of, or who died during the inpatient or observation hospital stay.

Exclusion Criteria:
Patients with inpatient LOS <24 hours are excluded because patients with brief hospital stays are less likely to have received the amount of medical intervention necessary to evaluate the quality of care.

Patients =18 years of age at admission are excluded because the Center of Excellence for Pediatric Quality Measurement's (CEPQM) task was to create a tool for measuring patient safety in the pediatric age group (i.e., <18 years of age). With this in mind, GAPPS is designed to perform exclusively in pediatric patients.

Table 2 - National Field Test Patient Characteristics:

Characteristics		
Gender, No. (%)		
Female	1,698	(44.5%)
Male	2,058	(54.0%)
Age, No. (%)		
0 to <3 months	1,031	(27.0%)
3 months to < 3 years	917	(24.0%)
3 years to < 10 years	784	(20.6%)
10 years to < 18 years	1,052	(27.6%)
Race/Ethnicity, No. (%)		
Black	477	(12.5%)
Hispanic	421	(11.0%)
White	2,165	(56.8%)
Other	184	(4.8%)
Missing	567	(14.9%)

Chronic Condition Indicators*, No. (%)		
0 body system	2,003	(52.5%)
1 body system	1,090	(28.6%)
2 body systems	324	(8.5%)
more than 2 body systems	129	(3.4%)
Insurance, No. (%)		
Private	2,073	(54.4%)
Public	1,306	(34.2%)
Other	106	(2.8%)
Length of Stay in day, Median (IQR)	3	(2,5)

* Chronic Condition Indicators allow researchers to determine whether a diagnosis is a chronic condition based on ICD-9-CM codes, and if so, what specific body system(s) are affected by a chronic condition.

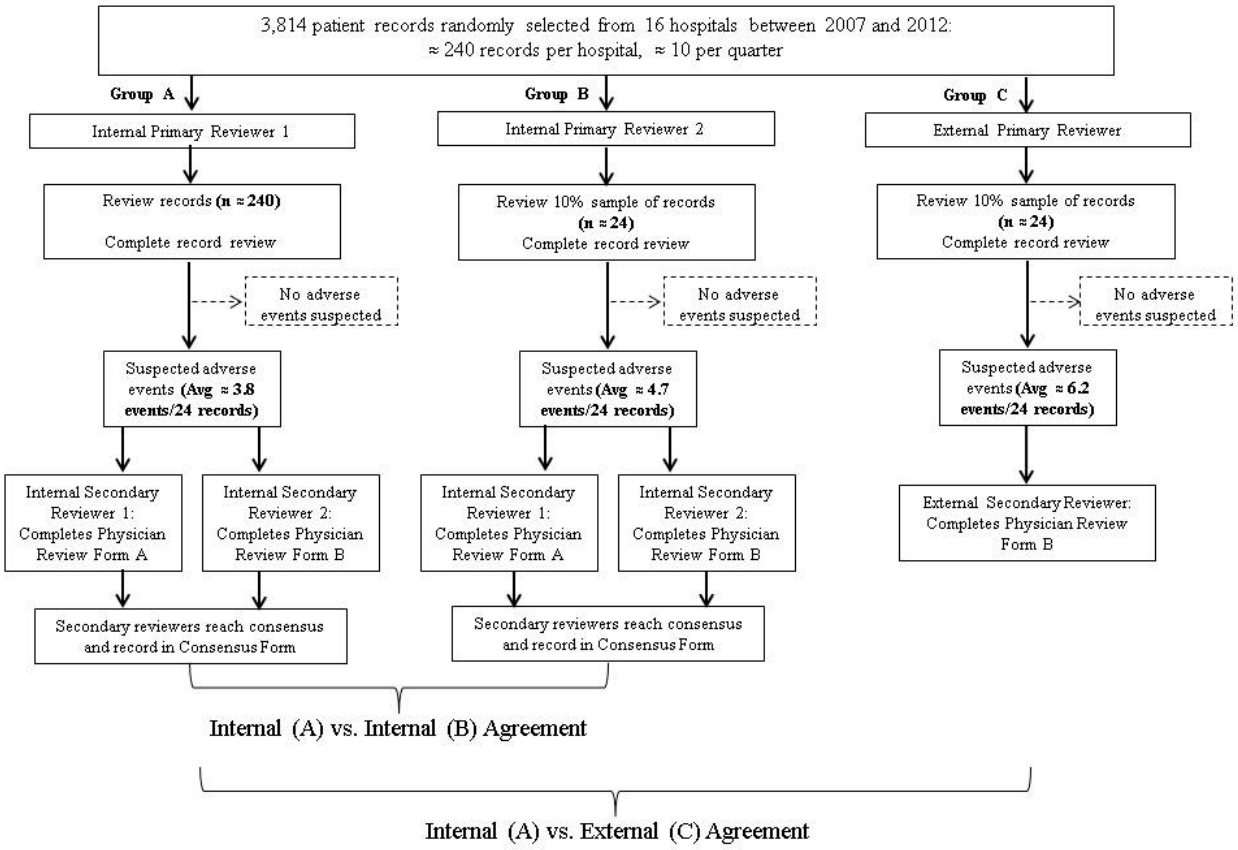
1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.
Not applicable.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
Race/Ethnicity
Presence of chronic conditions as measured by AHRQ's Chronic Condition Indicator (CCI) (38)
Insurance Status

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

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☐ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)
Figure 1 – Flowchart of the medical record review process in the GAPPS study:



The candidate triggers were evaluated using three groups of medical records in each hospital (see Figure 1). Group A consisted of the total sample of medical records from each hospital that were reviewed by primary (nurse) and secondary (physician) reviewers internal to the hospital. Group B consisted of a random sub-sample of the records in Group A (i.e., 24 records per site) that were reviewed again by additional primary and secondary reviewers internal to each hospital. Group C consisted of a separate random sub-sample (i.e., 24 records per site) of the records reviewed in Group A that were reviewed again by primary and secondary reviewers external to each hospital.

We used Groups A and B to evaluate the reliability of the measure. We compared ratings from pairs of independent secondary reviewers within Group A. We also compared primary reviewer findings (Group A) to a second primary reviewer's findings (Group B) for the same medical records. To assess reliability, we used a Kappa statistic for variables with only two possible outcomes and a weighted Kappa computed with Fleiss-Cohen weights for variables with more than two possible ordinal outcomes.(1,2) We used the categorization of Landis and Koch to interpret reliability for ranges of Kappa scores (k <0: poor, k = 0.00-0.20: slight, k = 0.21-0.40: fair, k = 0.41-0.60: moderate, k = 0.61-0.80: substantial, k = 0.81-1.00: almost perfect).

REFERENCES

1. Fleiss J, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Meas. 1973;33:613–9.

2. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics. 1977 Mar;33(1):159.

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

Table 3 - Internal Primary Reviewer (Nurse) Reliability:

Primary Review (Nurse)	Kappa	Agreement

# Triggers	0.68	83.1%
Any Trigger	0.67	87.9%
# Adverse Events	0.73	88.4%
Any Adverse Events	0.69	91.8%

Table 4 - Internal Secondary Reviewer (Physician) Reliability:

Secondary review (Physician)	Kappa	Agreement
Verification of Same Adverse Events	0.81	91.6%
AE Severity	0.86	89.6%
AE Preventability	0.72	77.4%

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

The reliability for internal primary reviewers (Group A versus Group B, [n = 379]) was "substantial" for both determination of the total number of suspected AEs (Kappa = 0.73, 95% CI 0.62 – 0.85) and identification of at least one suspected AE (Kappa = 0.69, 95% CI 0.59 – 0.79). Both primary reviewers agreed on the total number of AEs 88% of the time and agreed that a record did or did not contain at least one AE 92% of the time. In some cases, however, the AEs identified in the medical records differed. From the total sample of records reviewed, primary reviewers identified the same AEs 62% of the time.

The two internal secondary reviewers in Group A independently determined the presence or absence of an AE among suspected AEs identified by the primary reviewer in Group A (n = 617). Internal secondary reviewers verified the same suspected AEs 92% of the time, with "almost perfect" reliability (Kappa = 0.81, 95% CI 0.76 – 0.86).

The reliability for internal secondary reviewers in Group A versus Group B (n = 379) was "substantial" for both determination of the total number of suspected AEs (Kappa = 0.73, 95% CI 0.57 – 0.89) and verification of at least one suspected AE (Kappa = 0.70, 95% CI 0.59 – 0.81). Internal secondary reviewers in Group A and Group B agreed on the total number of AEs 92% of the time and agreed that a record did or did not contain at least one AE 94% of the time. In some cases, however, the AEs identified in the medical records differed.

As our team conducted reliability testing with patient-level data elements (AEs identified in individual medical records) using appropriate methodology (Kappa statistic for variables with only two possible outcomes and a weighted Kappa computed with Fleiss-Cohen weights for variables with more than two possible ordinal outcomes), and as there was substantial or moderate agreement between primary and secondary reviewers in both determination of the total number of suspected AEs and identification of at least one suspected AE, we deem that there is moderate confidence in the measure's reliability as per the NQF Algorithm #2: Guidance for Evaluating Reliability.

2b2. VALIDITY TESTING

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

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

☐ **Performance measure score**

- ☐ **Empirical validity testing**
- ☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

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

We developed the draft trigger tool used in the GAPPS measure through the RAND/UCLA Appropriateness Method, which is a modified Delphi process.(1–3) We first compiled a set of 78 candidate triggers from a literature review of existing pediatric and adult trigger tools and input from trigger tool experts.(4–6) We then recruited nine panelists from national pediatric and patient safety organizations and asked them to rate separately the validity and feasibility of the candidate triggers on a nine-point scale (where 1 is the least valid/feasible and 9 is the most valid/feasible). A trigger was considered valid if it was judged to be reasonably likely to identify an underlying AE, indicating that harm potentially occurred. A trigger was considered feasible if it was judged likely to be accurately and consistently documented in either paper or electronic medical records as part of patient care at a wide range of hospitals, from smaller community sites to larger tertiary care centers. Applying the RAND/UCLA Appropriateness Method, we accepted triggers that had both median validity and feasibility ratings greater than or equal to seven. This approach resulted in inclusion of 54 of the initial 78 candidate triggers in the draft GAPPS trigger list.

It is not possible to assess the performance of the GAPPS measure against a true “gold standard” for detection of preventable AEs because such a gold standard does not yet exist. We therefore focused our validity testing on evaluation of how accurately and completely "typical reviewers" (i.e., clinicians who are trained in GAPPS methodology but not necessarily trigger tool experts) were able to identify preventable AEs using the measure as compared to expert reviewers. The expert reviewers had extensive experience with using trigger tools for preventable AE identification and consequently were most likely to identify preventable AEs accurately and completely. To evaluate the validity of the GAPPS measure, we assessed the performance of the National Field Test hospitals' internal reviewers relative to the performance of external expert reviewers in applying the measure (i.e., we compared findings of reviewers in Group A versus Group C, as shown in Figure 1). For this comparison, we calculated the specificity and sensitivity between reviewer groups.

REFERENCES

1. Fitch K, Bernstein S, Aguilar MD, Burnand B, LaCalle JR, Lázaro P, et al. The RAND/UCLA Appropriateness Method User’s Manual. Santa Monica, CA: RAND; 2001.

2. Brown B. DELPHI PROCESS: A Methodology Used for the Elicitation of Opinions of Experts. Rand Corp. 1968 Sep;1–14.

3. Sweidan M, Williamson M, Reeve JF, Harvey K, O’Neill JA, Schattner P, et al. Identification of features of electronic prescribing systems to support quality and safety in primary care using a modified Delphi process. BMC Med Inform Decis Mak. 2010 Apr 15;10(1):21.

4. Stockwell D, Bisarya H, Classen D, Kirkendall E, Landrigan C, Lemon V, et al. A trigger tool to detect harm in pediatric inpatient settings. Pediatrics. 2015;

5. Griffin FA, Resar RK. IHI Global Trigger Tool for Measuring Adverse Events (Second Edition). Institute for Healthcare Improvement; 2009. (IHI Innovation Series white paper).

6. Kirkendall ES, Kloppenborg E, Papp J, White D, Frese C, Hacker D, et al. Measuring adverse events and levels of harm in pediatric inpatients with the Global Trigger Tool. Pediatrics. 2012 Nov;130(5):e1206-1214.

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

Table 5:

Identifying a record with 1 or more AEs	Specificity	Sensitivity
Internal primary reviewers versus External expert primary reviewers	0.91	0.40
Internal secondary reviewers versus External expert secondary reviewers	0.95	0.33

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)
As summarized in Table 5, using the findings of the external reviewers as the standard of comparison, the specificity for identifying a record with one or more AEs was 0.91 for primary reviewers and 0.95 after taking secondary reviewer verification into account. The sensitivity was 0.40 for primary reviewers and 0.33 after taking secondary reviewer verification into account. The lower sensitivity is likely due in part to the novice reviewers' lack of experience with the tool and their inability to make up for their inexperience by increasing the amount of time they took to perform their review, given that there was a 30-minute time limit per record.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

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

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)
Not applicable.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)
Not applicable.

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

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☒ **Statistical risk model with** 3 **risk factors**
☒ **Stratification by** 2 **risk categories**
☐ **Other,** Click here to enter description

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.
For inter-hospital comparisons of preventable adverse event rates, case-mix adjustment models should be fit with mixed effects negative binomial regression. In each model, the dependent variable is the number of preventable adverse events with exposure time equal to length of stay (a random intercept at the hospital level) and the fixed effects are the three case-mix adjusters: patient age group, number of chronic conditions, and service type (medical vs. surgery).

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

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

One of the methodological issues associated with making comparisons across institutions is the need to adjust appropriately for case-mix differences. Case-mix refers to patient characteristics, such as demographic characteristics and health status, that may affect measures of outcomes or processes. Systematic effects of this sort create the potential for a population's rates to be higher or lower because of its characteristics, rather than because of the quality of care provided, making comparisons of unadjusted rates potentially misleading. The basic goal of adjusting for case-mix is to estimate how different institutions would be rated if they all provided care to comparable groups of patients in terms of case-mix variables.

To evaluate potential variables for case-mix adjustment of GAPPS rates, we evaluated: patient age, sex, number of chronic conditions as determined using the Chronic Condition Indicator (CCI), and service type based on data collected by the reviewers from the patient record.(1) We included a variable in our multivariate case-mix models if its bivariate association with preventable adverse events was $p<0.10$.

REFERENCES

1. Chronic Condition Indicator (CCI) for ICD-10-CM [Internet]. Healthcare Cost and Utilization Project (HCUP); Available from: https://www.hcup-us.ahrq.gov/toolssoftware/chronic_icd10/chronic_icd10.jsp

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

Gender was not associated with preventable adverse events (PAEs) in both the teaching and community hospitals ($p=0.69$ and $p=0.29$, respectively). Age group (< 3 years, 3 years to < 10 years, 10 years to < 18 years) was associated with PAEs in the community ($p<0.001$) and teaching ($p=0.01$) hospitals. Surgery services (versus medical) were also associated with PAEs in both the community ($p=0.002$) and the teaching ($p<0.001$) hospitals. We evaluated the association of the number of CCIs (0, 1, 2, and 3+) and PAEs and found that in community and teaching hospitals the risk associated with 0 and 1 CCIs was not different ($p=0.49$ and $p=0.90$, respectively), and the risk associated with 2 and 3+ CCIs was not different ($p=0.91$ and p-0.87, respectively). After collapsing the CCI variable, patients with 2+ CCIs were significantly more likely to have PAEs compared to patients with 0 or 1 CCIs in both community ($p=0.009$) and teaching ($p=0.008$) hospitals. Therefore, we included CCIs (2+ CCIs versus 0-1 CCI), age (< 3 years, 3 years to < 10 years, 10 years to < 18 years), and type of service (surgery versus medical) as adjustors in our case-mix model.

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

GAPPS is a measure of preventable adverse events in hospitals. It is measuring in-hospital processes of care (e.g., the safety of medication ordering and delivery, procedural performance, care coordination) that should be equally applied to all, regardless of SDS. Unlike many other common measures (e.g., readmissions, where social factors beyond the purview of the healthcare system are an important factor in readmission rates), GAPPS is focused on in-hospital care quality that is within the control of the healthcare system largely irrespective of patient SDS. Therefore we have chosen not to include SDS characteristics in our risk adjustment model because we do not believe that SDS characteristics should be inherently related to rates of preventable adverse events at a substantial level, except insofar as they may be associated with true differences in care quality. That said, we do recommend the reporting of stratified analyses by SDS groups, to facilitate identification of disparities in care.

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

Pseudo R-squared values for the teaching hospital and community hospital models were 0.02 and 0.11, respectively.

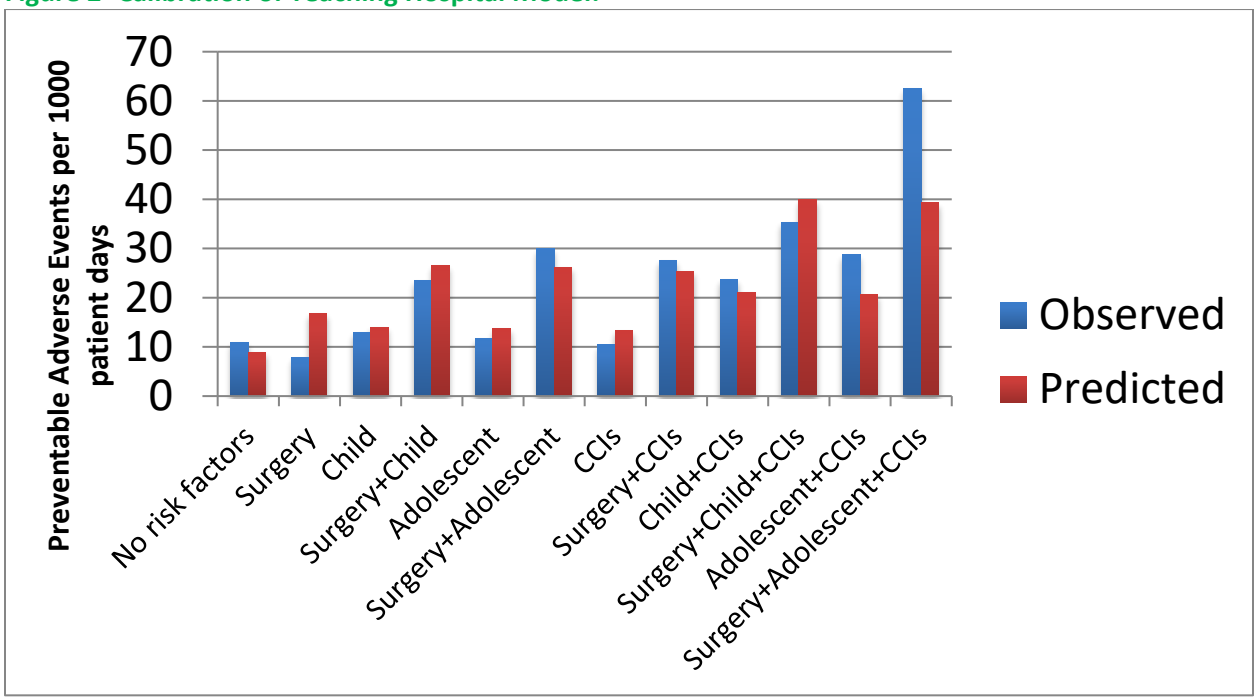
Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
Our models showed good discrimination between any PAEs and no PAEs (community hospitals c-statistic= 0.80, 95% confidence interval [CI] 0.69-0.91; teaching hospitals c-statistic= 0.76, 95% CI 0.71-0.81).

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
The teaching and community hospital models had good calibration across all twelve risk groups represented in our model (goodness-of-fit tests p=0.31 and p=0.99, respectively).

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Figure 2- Calibration of Teaching Hospital Model:



2b4.9. Results of Risk Stratification Analysis:
Table 6 - Model results for teaching and community hospitals:

	Teaching Hospitals		
	Coefficient	95% confidence interval	p-value
Number of CCIs (0 or 1=reference)			
2+ CCIs	0.410	0.052, 0.767	0.02
Age group (< 3 years = reference			
≥3 years and <10 years	0.458	0.016, 0.899	0.04
≥10 years and <18 years	0.439	0.056, 0.822	0.02
Service type (medical = reference)			
Surgery	0.638	0.225, 1.051	0.002
β_0	-4.724	-5.000, -4.448	<0.001

	Community Hospitals		
	Coefficient	95% confidence interval	p-value

Number of CCIs (0 or 1=reference)			
2+ CCIs	1.406	-0.055, 1.867	0.06
Age group (< 3 years = reference			
≥3 years and <10 years	1.450	-0.062, 2.962	0.06
≥10 years and <18 years	2.041	0.704, 3,378	0.003
Service type (medical = reference)			
Surgery	0.676	-0.652, 2.005	0.32
β_0	-7.060	-7.975, -6,145	<0.001

Based on the above models, we recommend stratification by hospital type. Additionally, we recommend use of the following three categorical variables in the GAPPS case-mix adjustment model: age, number of CCIs, and service type.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)
Our models showed good discrimination and were well calibrated across risk groups.

In many analyses, the goal is to explain as much of the variance as possible, in which case a high R-squared is desired. In this case, the value of the R-squared represents the extent to which case-mix adjustment affected measure scores. For example, if the case-mix adjusters had no effect (e.g., age was not predictive of measure scores), then the R-squared value would be zero. Overall, case-mix adjustments explained a small proportion of the variation in preventable adverse events.

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

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*
Since GAPPS facilitates preventable AE identification among pediatric inpatients, teams can conduct intra-hospital comparisons of preventable AE rates between different hospital divisions. Hospitals can also track their preventable AE rates over time to evaluate their state of patient safety, and can compare preventable AEs based on preventability and severity.

Comparison of preventable AE rates across hospitals would require reviewers at each institution to receive adequate training in the trigger tool methodology to ensure standardization of the preventable AE detection process. We anticipate that such training would contribute to increasing the measure's reliability, which would be necessary to reach a level appropriate for inter-hospital comparisons. Using the automated approach would further increase reliability because it removes the human error involved in finding triggers. However, using an automated trigger identification system instead of the manual system has no impact on the rest of the measure calculation, and we anticipate there are no obvious differences for either approach. A national, state, or other multi-hospital database would be ideal for inter-hospital comparisons but does not yet exist. If GAPPS is used for inter-hospital comparisons, preventable AE rates can be compared based on hospital type (academic vs. community, see figure 3).

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

Figure 3 - Distribution of preventable AE rates by hospital:

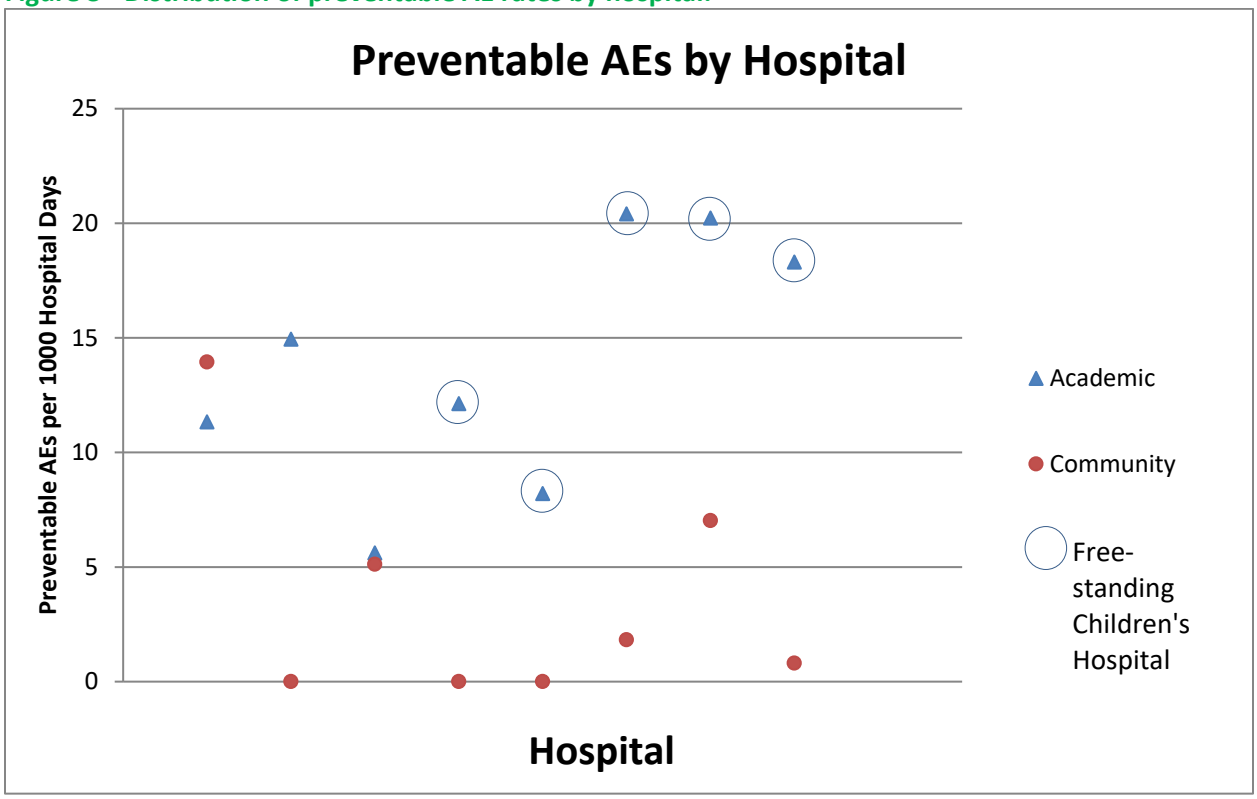
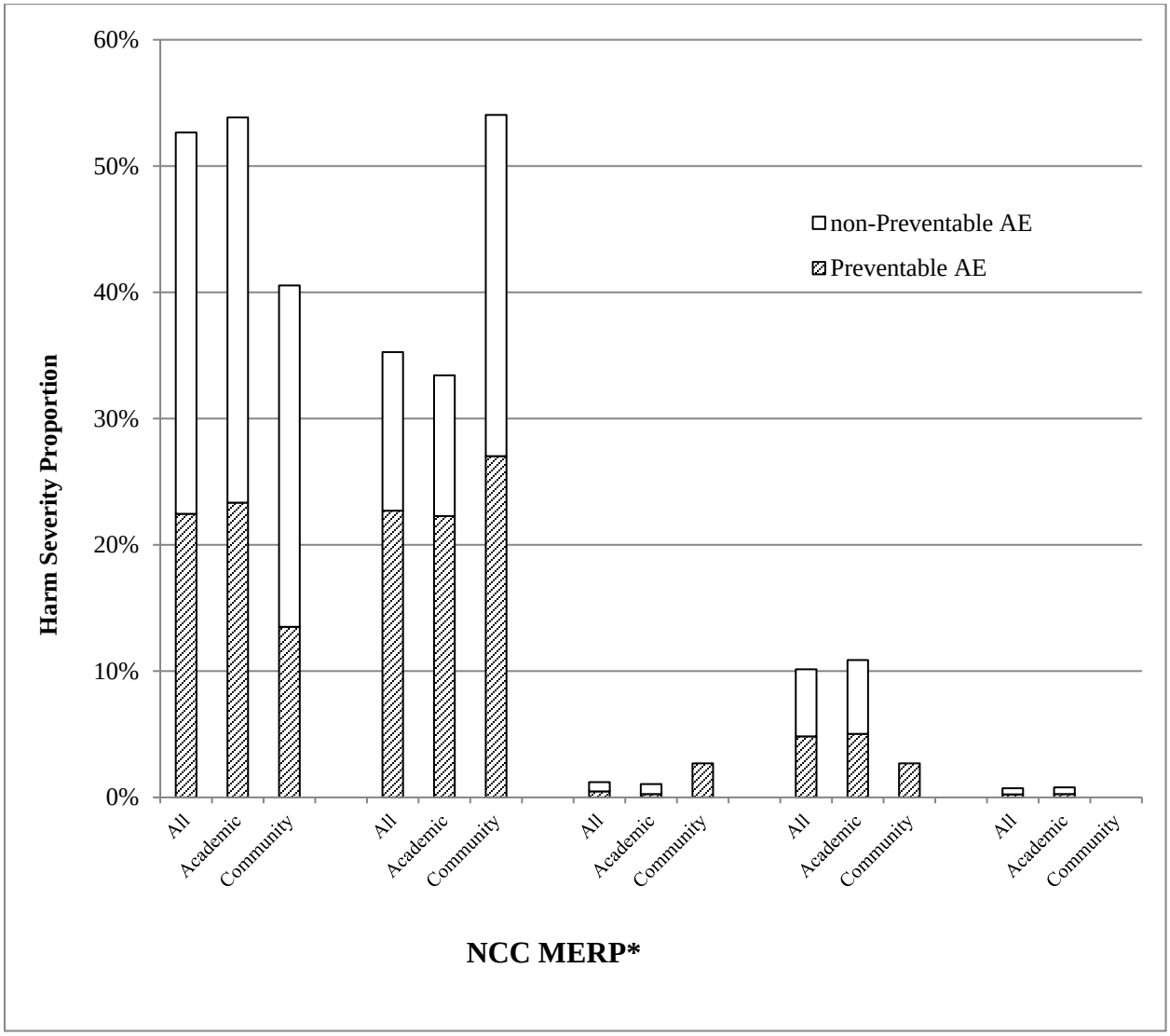
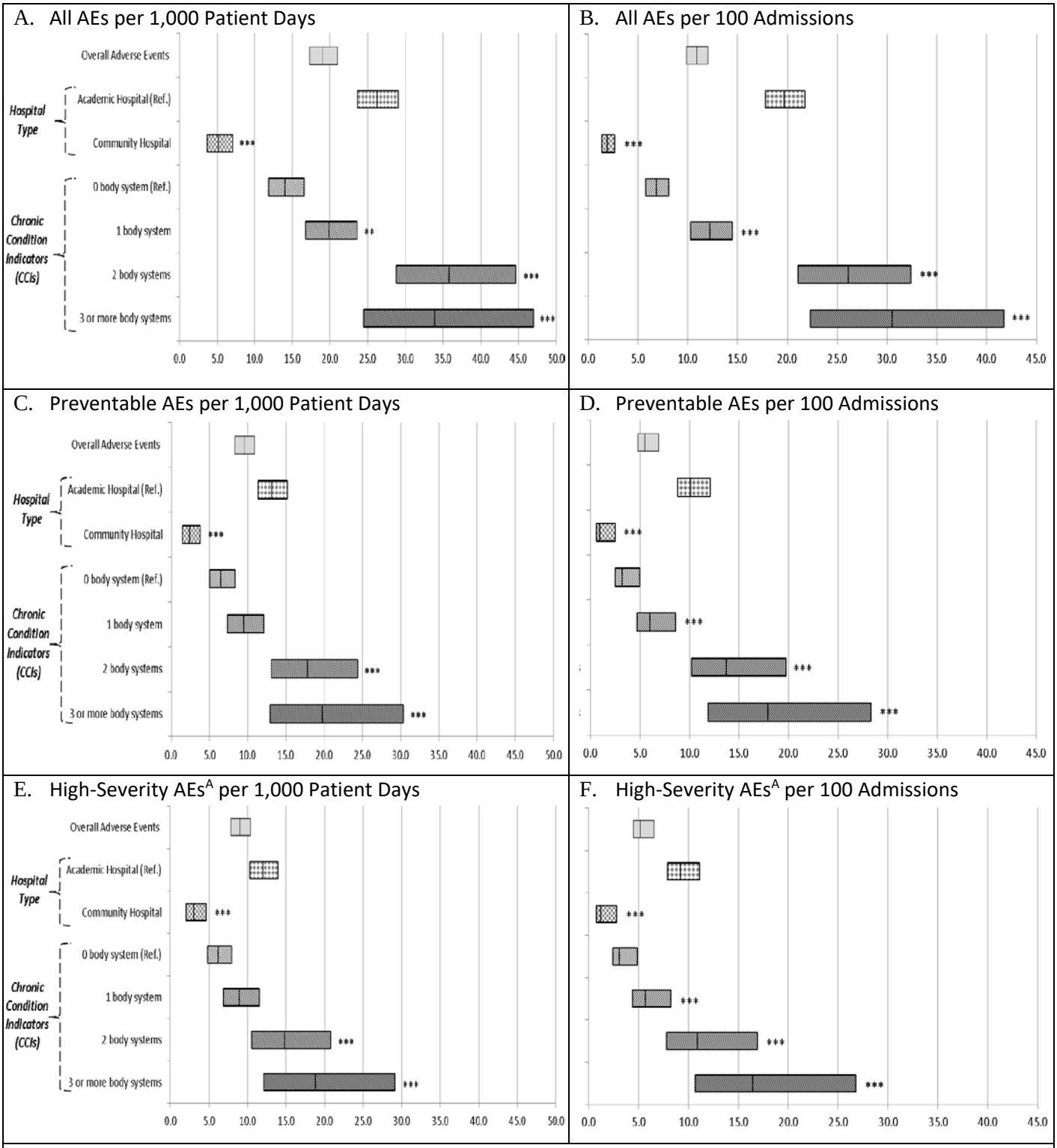


Figure 4 - Severity of all harms and preventable harms:



* NCC MERP Categories: E: contributed to or resulted in temporary harm to the patient and required intervention, F: contributed to or resulted in temporary harm to the patient and required initial or prolonged hospitalization, G: contributed to or resulted in permanent patient harm, H: required intervention to sustain life, I: contributed to or resulted in the patient's death.

Figure 5 - Distribution of Adverse Events (AEs) by Hospital and Clinical Characteristics:

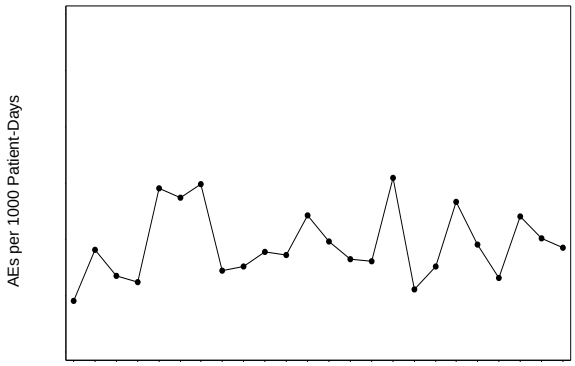


A: High-Severity AEs are defined as NCC MERP categories F to I

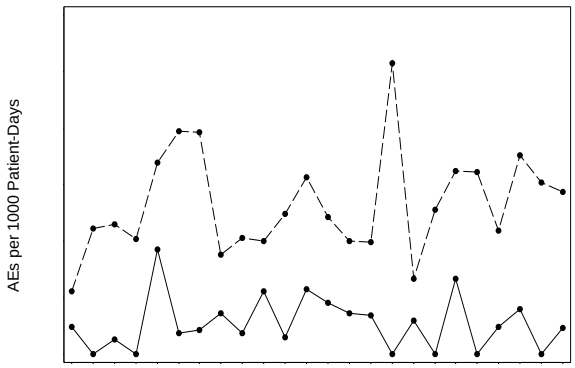
**: p<0.01; ***: p<0.001

Figure 6 – Rates of all harms, preventable harms, and high-severity harms per 1,000 patient-days, according to quarter:

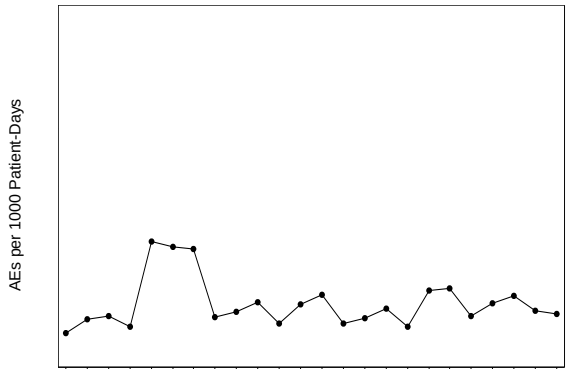
A. All AEs



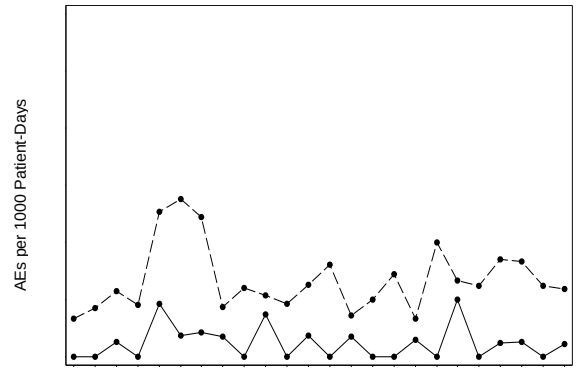
B. By Hospital Type[A], All AEs



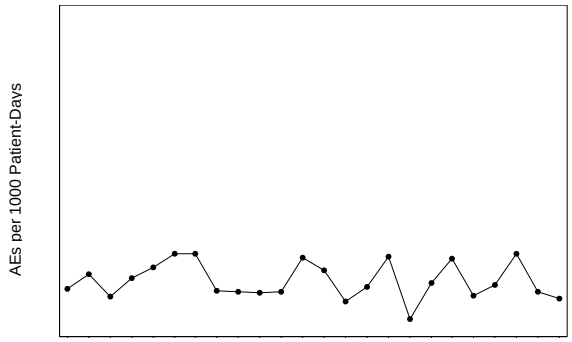
C. Preventable AEs



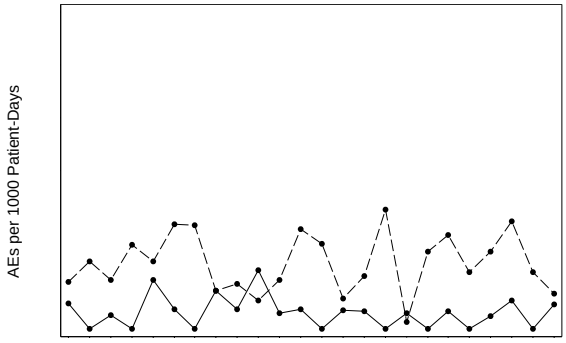
D. By Hospital Type[A], Preventable AEs



E. High-Severity AEs[B]

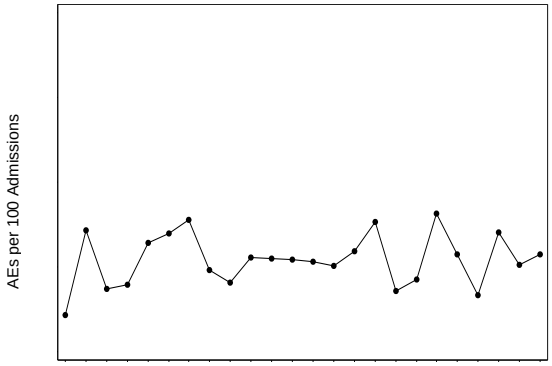


F. By Hospital Type[A], High-Severity AEs

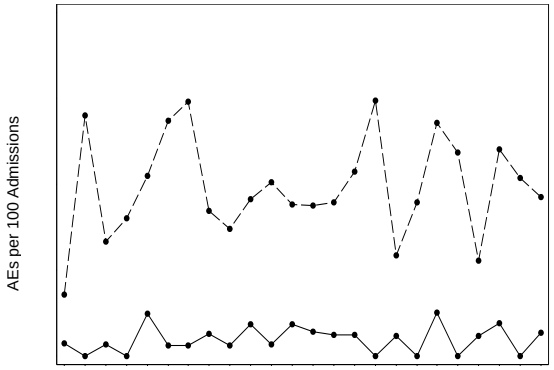


A: ------- Academic Hospital, —— Community Hospital
B: High-severity AEs are defined as NCC MERP categories F to I
Note: The values presented are unadjusted.

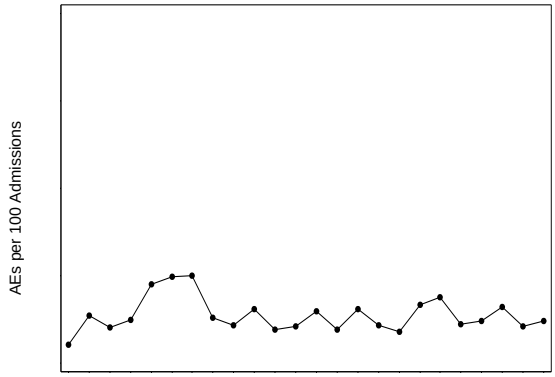
A. All AEs



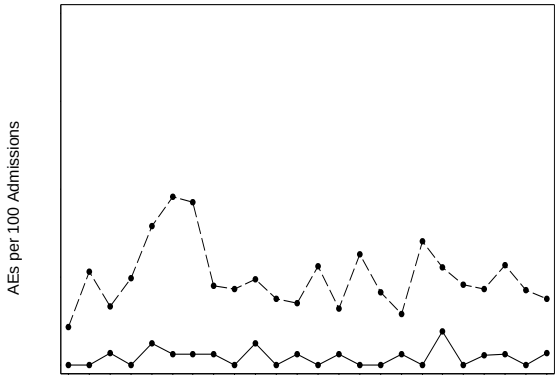
B. By Hospital Type[A], All AEs



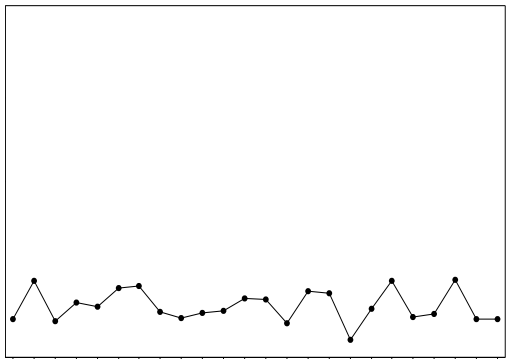
C. Preventable AEs



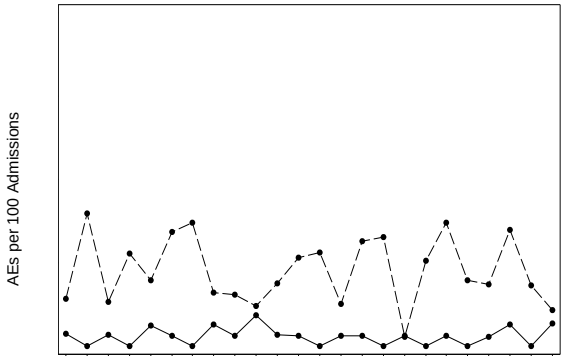
D. By Hospital Type[A], Preventable AEs



E. High-Severity AEs[B]



F. By Hospital Type[A], High-Severity AEs



A: ------- Academic Hospital, —— Community Hospital
B: High-severity AEs are defined as NCC MERP categories F to I
Note: The values presented are unadjusted.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do*

the results mean in terms of statistical and meaningful differences?)
In a study of 16 academic and community hospitals that care for children across the US, we found that harm due to medical care remained common from 2007 to 2012 and did not decrease significantly over time. Approximately half of all harms are preventable, indicating that these harms can be immediately targeted by quality improvement teams. AEs were most commonly severity levels of E and F.

We found wide disparities in the rates of harm in academic and community hospitals. The reasons for this difference are unclear, but major differences in the frequency of complex chronic conditions as well as in the types and severity of illness seen in the two types of hospitals likely explain much of the difference. Neither community nor academic centers experienced improvements over the six-year span studied, suggesting that effectively controlling pediatric patient safety problems has proven similarly difficult in both settings, despite their baseline differences in populations and harm epidemiology.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

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

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

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
The only source of missing data for this measure is missing documents for a selected hospitalization. In the rare circumstance that a medical record was missing all documents, the selected hospitalization was discarded and replaced by another hospitalization. We only have record of this happening with two medical records in our study (n=3,814). We have no reason to believe that missing a record is anything more than a random and rare, non-systematic event. With the increasing uptake of electronic medical records and transition away from paper charts, missing documentation for hospitalizations will become vanishingly rare. For these reasons, we believe missing data did not represent a significant

source of systematic bias. Given the infrequency of missing data, we were unable to conduct additional testing on the cause or impact of missing documentation.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)
Given the rarity of missing documents, especially in settings which have adopted electronic health records, no overall frequency of missing data was calculated. As a result, no additional testing was conducted related to the missing data. In addition, we do not believe missing data could have biased results based on the infrequency of missing records.

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

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
If other:

3b. Electronic Sources

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

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

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).
The measure currently requires a clinician to review each adverse event in order to determine whether that event may have been preventable. It may be possible in the future for this step to transition to an automated process. While capturing triggers can be automated, a completely automated approach is currently not feasible and also not likely feasible for the near future.

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

3c. Data Collection Strategy

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

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.
Not Applicable

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
The codes for the GAPPS automated triggers, the GAPPS Manual of Operations, and all associated forms that reviewers complete are available to users free of charge.

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting Quality Improvement (external benchmarking to organizations) Quality Improvement (Internal to the specific organization)	

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

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

Not applicable.

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

GAPPS was newly commissioned and developed as part of the AHRQ/CMS Pediatric Quality Measures Program and is therefore not yet in use or publicly reported.

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

AHRQ and CMS intend that the GAPPS measure be available for public use with the current expectation that the full measure specifications be provided on the AHRQ website, CMS website, or both. For ease of implementation, we have prepared the GAPPS Manual of Operations and automated trigger codes for detection and analysis of preventable AE rates (see Appendix A). Our testing has shown that the measure is straightforward to implement across a variety of hospital types and on both paper and electronic medical records. In addition, we have made a series of comprehensive training videos that are easily accessible online for sites that want to learn how to utilize GAPPS.

Although GAPPS is not currently used for public reporting, endorsement will facilitate the measure's use by public and private payers, provider organizations, and consumer groups that require NQF endorsement of quality measures and will help support the integration of GAPPS into other patient safety measures. We anticipate that GAPPS results will be useful to everyone with a need for information on the quality of pediatric inpatient care, including patients, parents, hospitals, health plans, insurers, and policy makers. In addition, hospitals could provide GAPPS performance scores to quality organizations and purchasers. GAPPS reliably identifies preventable AEs and can be used to guide and monitor quality improvement efforts and facilitate inter-hospital comparisons.

Improvement

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

4b. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

POTENTIAL FOR QUALITY IMPROVEMENT

Key strategies for reducing preventable harms in children include early detection and treatment of potential harm(1) and identification of potentially preventable adverse events.(2) Use and further development of measures such as GAPPS to detect adverse events is thus a critical part of efforts to improve patient safety.(3–5) By using more sensitive and reliable measures, hospitals can increase their capacity to quantify inpatient adverse events, identify priorities, and target available resources.(6) Multiple studies have shown that hospitals with reliable means to track adverse events have experienced improvements in patient safety and associated clinical outcomes.(7–12)

REFERENCES

1. Larsen GY, Donaldson AE, Parker HB, Grant MJCP. Preventable harm occurring to critically ill children*. Pediatr Crit Care Med July 2007. 2007;8(4):331–6.
2. McDonald KM, Davies SM, Haberland CA, Geppert JJ, Ku A, Romano PS. Preliminary Assessment of Pediatric Health Care Quality and Patient Safety in the United States Using Readily Available Administrative Data. Pediatrics. 2008 Aug 1;122(2):e416–25.
3. Takata GS, Mason W, Taketomo C, Logsdon T, Sharek PJ. Development, testing, and findings of a pediatric-focused trigger tool to identify medication-related harm in US children's hospitals. Pediatrics. 2008 Apr;121(4):e927-935.
4. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. CMAJ Can Med Assoc J J Assoc Medicale Can. 2004 May 25;170(11):1678–86.
5. Agarwal S, Classen D, Larsen G, Tofil NM, Hayes LW, Sullivan JE, et al. Prevalence of adverse events in pediatric intensive care units in the United States. Pediatr Crit Care Med J Soc Crit Care Med World Fed Pediatr Intensive Crit Care Soc. 2010 Sep;11(5):568–78.
6. Miller MR, Elixhauser A, Zhan C. Patient Safety Events During Pediatric Hospitalizations. Pediatrics. 2003 Jun 1;111(6):1358–66.
7. Evans RS, Pestotnik SL, Classen DC, Bass SB, Burke JP. Prevention of adverse drug events through computerized surveillance. Proc Annu Symp Comput Appl Sic Med Care Symp Comput Appl Med Care. 1992;437–41.
8. Kaufman J, Rannie M, Kahn MG, Vitaska M, Wathen B, Peyton C, et al. An interdisciplinary initiative to reduce unplanned extubations in pediatric critical care units. Pediatrics. 2012 Jun;129(6):e1594-1600.
9. Piontek F, Kohli R, Conlon P, Ellis JJ, Jablonski J, Kini N. Effects of an adverse-drug-event alert system on cost and quality outcomes in community hospitals. Am J Health-Syst Pharm AJHP Off J Am Soc Health-Syst Pharm. 2010 Apr 15;67(8):613–20.
10. Cohen MM, Kimmel NL, Benage MK, Cox MJ, Sanders N, Spence D, et al. Medication safety program reduces adverse drug events in a community hospital. Qual Saf Health Care. 2005 Jun;14(3):169–74.
11. Garrett PR Jr, Sammer C, Nelson A, Paisley KA, Jones C, Shapiro E, et al. Developing and implementing a standardized process for global trigger tool application across a large health system. Jt Comm J Qual Patient Saf Jt Comm Resour. 2013 Jul;39(7):292–7.
12. Miles A, Spaeder M, Stockwell D. Unplanned ICU Transfers from Inpatient Units: Examining the Prevalence and Preventability of Adverse Events Associated with ICU Transfer in Pediatrics. J Pediatr Intensive Care. 2015 Nov 21;5(1):021–7.

4c. Unintended Consequences

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

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

No unexpected findings were identified during testing.

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

No unexpected benefits were identified during testing.

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
During development of the original trigger list, we conducted a nine-member expert panel from top national stakeholder organizations using the RAND/UCLA Appropriateness Method. Their feedback was critical for determining which triggers to include in our final trigger list and for moving the project from development to field testing. Some of the site leads participating in the National Field Test (NFT) provided feedback on the development of the NFT materials and procedures. Following the NFT, results, data, and interpretation assistance were provided to the participating institutions.

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
Results reports were created after development and field testing of the measure. The reports included field test summary results, brief explanations of these results, and the raw institution-specific data. These reports were sent to the NFT institutions that requested their results following completion of the field test and analysis of the data.

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.
Feedback was obtained through personal correspondence with the GAPPS team. Trained reviewers from the participating NFT sites reported understanding GAPPS and were able to identify triggers in medical records, use them to detect AEs, and assess severity and preventability. Feedback from the site leads indicated that the measure is straightforward to use and easily understandable. Our results indicated that GAPPS works for both EHRs and paper medical records.

4d2.2. **Summarize the feedback obtained from those being measured.**
Please see above.

4d2.3. **Summarize the feedback obtained from other users**
Throughout development and testing, the GAPPS team presented our candidate measure to our Scientific Advisory Board, consisting of representatives from Boston Children´s Hospital, the larger Harvard community, and organizations such as the National Initiative for Children´s Healthcare Quality, as well as to our National Stakeholder Panel, which includes representatives from diverse national organizations that represent patients and families, providers, payers, and health services researchers. Comments and feedback from these various users and stakeholders indicate that they believe such a tool is useful to pediatric medical settings. Since finalizing the GAPPS measure, the team has presented their findings at national conferences and other public forums. We have received positive feedback and many stated that the measure would be useful to them.

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
Based on helpful suggestions from the NFT sites and reviewers, we improved the clarity of the Manual of Operations. The GAPPS NFT also demonstrated that it is crucial to provide rigorous training and feedback to reviewers on practice cases prior to reviewers' use of the measure in order to achieve optimal standardization in AE detection.

5. Comparison to Related or Competing Measures

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

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

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

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

5a. Harmonization of Related Measures

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

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

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

5b. Competing Measures

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

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

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)
If other:

3b. Electronic Sources

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

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

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

The measure currently requires a clinician to review each adverse event in order to determine whether that event may have been preventable. It may be possible in the future for this step to transition to an automated process. While capturing triggers can be automated, a completely automated approach is currently not feasible and also not likely feasible for the near future.

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

3c. Data Collection Strategy

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

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.
Not Applicable

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
The codes for the GAPPS automated triggers, the GAPPS Manual of Operations, and all associated forms that reviewers complete are available to users free of charge.

4. Usability and Use

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

4a. Accountability and Transparency

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

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting Quality Improvement (external benchmarking to organizations) Quality Improvement (Internal to the specific organization)	

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

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

Not applicable.

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
GAPPS was newly commissioned and developed as part of the AHRQ/CMS Pediatric Quality Measures Program and is therefore not yet in use or publicly reported.

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
AHRQ and CMS intend that the GAPPS measure be available for public use with the current expectation that the full measure specifications be provided on the AHRQ website, CMS website, or both. For ease of implementation, we have prepared the GAPPS Manual of Operations and automated trigger codes for detection and analysis of preventable AE rates (see Appendix A). Our testing has shown that the measure is straightforward to implement across a variety of hospital types and on both paper and electronic medical records. In addition, we have made a series of comprehensive training videos that are easily accessible online for sites that want to learn how to utilize GAPPS.

Although GAPPS is not currently used for public reporting, endorsement will facilitate the measure's use by public and private payers, provider organizations, and consumer groups that require NQF endorsement of quality measures and will help support the integration of GAPPS into other patient safety measures. We anticipate that GAPPS results will be useful to everyone with a need for information on the quality of pediatric inpatient care, including patients, parents, hospitals, health plans, insurers, and policy makers. In addition, hospitals could provide GAPPS performance scores to quality organizations and purchasers. GAPPS reliably identifies preventable AEs and can be used to guide and monitor quality improvement efforts and facilitate inter-hospital comparisons.

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

4b. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
POTENTIAL FOR QUALITY IMPROVEMENT
Key strategies for reducing preventable harms in children include early detection and treatment of potential harm(1) and identification of potentially preventable adverse events.(2) Use and further development of measures such as GAPPS to detect adverse events is thus a critical part of efforts to improve patient safety.(3–5) By using more sensitive and reliable measures, hospitals can increase their capacity to quantify inpatient adverse events, identify priorities, and target available resources.(6) Multiple studies have shown that hospitals with reliable means to track adverse events have experienced improvements in patient safety and associated clinical outcomes.(7–12)

REFERENCES
1. Larsen GY, Donaldson AE, Parker HB, Grant MJCP. Preventable harm occurring to critically ill children*. Pediatr Crit Care Med July 2007. 2007;8(4):331–6.
2. McDonald KM, Davies SM, Haberland CA, Geppert JJ, Ku A, Romano PS. Preliminary Assessment of Pediatric Health Care Quality and Patient Safety in the United States Using Readily Available Administrative Data. Pediatrics. 2008 Aug 1;122(2):e416–25.
3. Takata GS, Mason W, Taketomo C, Logsdon T, Sharek PJ. Development, testing, and findings of a pediatric-focused trigger tool to identify medication-related harm in US children's hospitals. Pediatrics. 2008 Apr;121(4):e927-935.
4. Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. CMAJ Can Med Assoc J J Assoc Medicale Can. 2004 May 25;170(11):1678–86.

5. Agarwal S, Classen D, Larsen G, Tofil NM, Hayes LW, Sullivan JE, et al. Prevalence of adverse events in pediatric intensive care units in the United States. Pediatr Crit Care Med J Soc Crit Care Med World Fed Pediatr Intensive Crit Care Soc. 2010 Sep;11(5):568–78.
6. Miller MR, Elixhauser A, Zhan C. Patient Safety Events During Pediatric Hospitalizations. Pediatrics. 2003 Jun 1;111(6):1358–66.
7. Evans RS, Pestotnik SL, Classen DC, Bass SB, Burke JP. Prevention of adverse drug events through computerized surveillance. Proc Annu Symp Comput Appl Sic Med Care Symp Comput Appl Med Care. 1992;437–41.
8. Kaufman J, Rannie M, Kahn MG, Vitaska M, Wathen B, Peyton C, et al. An interdisciplinary initiative to reduce unplanned extubations in pediatric critical care units. Pediatrics. 2012 Jun;129(6):e1594-1600.
9. Piontek F, Kohli R, Conlon P, Ellis JJ, Jablonski J, Kini N. Effects of an adverse-drug-event alert system on cost and quality outcomes in community hospitals. Am J Health-Syst Pharm AJHP Off J Am Soc Health-Syst Pharm. 2010 Apr 15;67(8):613–20.
10. Cohen MM, Kimmel NL, Benage MK, Cox MJ, Sanders N, Spence D, et al. Medication safety program reduces adverse drug events in a community hospital. Qual Saf Health Care. 2005 Jun;14(3):169–74.
11. Garrett PR Jr, Sammer C, Nelson A, Paisley KA, Jones C, Shapiro E, et al. Developing and implementing a standardized process for global trigger tool application across a large health system. Jt Comm J Qual Patient Saf Jt Comm Resour. 2013 Jul;39(7):292–7.
12. Miles A, Spaeder M, Stockwell D. Unplanned ICU Transfers from Inpatient Units: Examining the Prevalence and Preventability of Adverse Events Associated with ICU Transfer in Pediatrics. J Pediatr Intensive Care. 2015 Nov 21;5(1):021–7.

4c. Unintended Consequences

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

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

No unexpected findings were identified during testing.

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

No unexpected benefits were identified during testing.

4d1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

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

During development of the original trigger list, we conducted a nine-member expert panel from top national stakeholder organizations using the RAND/UCLA Appropriateness Method. Their feedback was critical for determining which triggers to include in our final trigger list and for moving the project from development to field testing. Some of the site leads participating in the National Field Test (NFT) provided feedback on the development of the NFT materials and procedures. Following the NFT, results, data, and interpretation assistance were provided to the participating institutions.

4d1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

Results reports were created after development and field testing of the measure. The reports included field test summary results, brief explanations of these results, and the raw institution-specific data. These reports were sent to the NFT institutions that requested their results following completion of the field test and analysis of the data.

4d2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

Feedback was obtained through personal correspondence with the GAPPS team. Trained reviewers from the participating NFT sites reported understanding GAPPS and were able to identify triggers in medical records, use them to detect AEs, and assess severity and preventability. Feedback from the site leads indicated that the measure is straightforward to use and easily understandable. Our results indicated that GAPPS works for both EHRs and paper medical records.

4d2.2. Summarize the feedback obtained from those being measured.

Please see above.

4d2.3. **Summarize the feedback obtained from other users**
Throughout development and testing, the GAPPS team presented our candidate measure to our Scientific Advisory Board, consisting of representatives from Boston Children´s Hospital, the larger Harvard community, and organizations such as the National Initiative for Children´s Healthcare Quality, as well as to our National Stakeholder Panel, which includes representatives from diverse national organizations that represent patients and families, providers, payers, and health services researchers. Comments and feedback from these various users and stakeholders indicate that they believe such a tool is useful to pediatric medical settings. Since finalizing the GAPPS measure, the team has presented their findings at national conferences and other public forums. We have received positive feedback and many stated that the measure would be useful to them.

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
Based on helpful suggestions from the NFT sites and reviewers, we improved the clarity of the Manual of Operations. The GAPPS NFT also demonstrated that it is crucial to provide rigorous training and feedback to reviewers on practice cases prior to reviewers' use of the measure in order to achieve optimal standardization in AE detection.

5. Comparison to Related or Competing Measures

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

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

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

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

5a. Harmonization of Related Measures
The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

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

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

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

Not Applicable

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** AAA_MOO_-_Appendices_FINAL_20170109.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Center of Excellence for Pediatric Quality Measurement
Co.2 **Point of Contact:** Mark, Schuster, cepqm@childrens.harvard.edu, 617-355-5859-
Co.3 **Measure Developer if different from Measure Steward:** Center of Excellence for Pediatric Quality Measurement
Co.4 **Point of Contact:** Mark, Schuster, cepqm@childrens.harvard.edu, 617-355-5859-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
The following people participated in measure development:

CORE TEAM:
Mark A. Schuster, MD, PhD (Principal Investigator, Director)
Christopher P. Landrigan, MD, MPH (Measure Co-Leader)
David C. Stockwell, MD, MBA (Measure Co-Leader)
Sara L. Toomey, MD, MPhil, MPH, MSc (Managing Director)

ADDITIONAL MEMBERS OF THE GAPPS TEAM, ESPECIALLY:
Jisun Jang, MA
Jessica A. Quinn, MS

ADDITIONAL CONTRIBUTORS:
Kimberly A. Chin, BA
Shannon M. Cottreau, BSN, RN, CPN
Colleen T. Madden, RN, BSN
Mari M. Nakamura, MD, MPH
Gareth J. Parry, PhD

RESEARCH ASSISTANTS:
Sepideh Ashrafzadeh, BS
Samuel S. Loren, BS
Sifon U. Ndon, AB
Debanjan Pain, AB
Michaela S. Tracy, BA
Michelle J. Wang, AB
Melody Wu, AB

Members of the GAPPS Advisory Committee: Hema Bisarya, MHSA, RD, David C. Classen, MD, MS, Paul J. Sharek, MD, MPH, and Rajendu Srivastava, MD, MPH

External expert primary reviewers: Kathleen M. Haig, RN, Diedre A. Rahn, RN, and Katherine R. Zigmont, RN

Expert secondary reviewers and training video instructors: Lee M. Adler, DO and Roger K. Resar, MD

Members of the GAPPS Expert Stakeholder Panel: David Bundy, MD, MPH, S. Todd Callahan, MD, MPH, Emi Datuin-Pal, RN, BSN, MSHSA, MBA, Carol Haraden, PhD, Laura Knobel, MD, FAAFP, Rita Pickler, PhD, RN, PNP-BC, FAAN, Xavier Sevilla, MD, MBA, FAAP, Jennifer Slayton, MSN, RN, and Glenn Takata, MD, MS

Reviewers and leads at all of our participating study sites: Boston Children's Hospital, Children's Hospital Colorado, Children's National Medical Center, Cincinnati Children's Hospital Medical Center, Grand View Hospital, Mary Washington Hospital, Lucile Packard Children's Hospital Stanford, Providence St. Peter Hospital, Progress West Hospital, University of Florida Health Shands Children's Hospital, Silver Cross Hospital, New York Presbyterian/Weill Cornell Medical Center, Utah Valley Regional Medical Center, Western Virginia University Hospitals, Hillcrest Hospital, and South Shore Hospital

Staff of the Center of Excellence for Pediatric Quality Measurement (CEPQM) at Boston Children's Hospital, members of CEPQM's Scientific Advisory Board, members of the National Stakeholder Panel Members of the Massachusetts Child Health Quality Coalition, and the GAPPS Expert Stakeholder Panel provided guidance and feedback on the measure.

Measure Developer/Steward Updates and Ongoing Maintenance

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

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

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

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

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:**

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

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3153
Corresponding Measures:
De.2. **Measure Title:** Continuity of Primary Care for Children with Medical Complexity
Co.1.1. **Measure Steward:** Seattle Children´s Research Institute
De.3. **Brief Description of Measure:** This measure assesses the percentage of children with medical complexity age 1 to 17 years old who have a Bice-Boxerman continuity of care index (hereafter referred to as Bice-Boxerman COC index) of >=0.5 in the primary care setting over a 12-month period.
1b.1. **Developer Rationale** Increasing numbers of children in the United States are living with medical complexity. (1) An analysis of the 2005-2006 National Survey of Children with Special Health Care Needs found that children in the highest category of complexity represent 0.4% of all children in the United States. In a 2014 study in 12 states, children with medical complexity (CMC) represented only 5.8% of children covered by Medicaid but accounted for 34% of all healthcare expenditures.(2) Given the challenges of coordinating care for these children, a continuous relationship with a single primary care provider or small number of providers with in-depth knowledge of their medical and social needs is essential.

CMC stand to benefit from continuity of care because repeated contact could provide more opportunities to discuss their needs and receive help coordinating the many providers and services involved in their care outside of the primary care setting. Nationally, families of children with special needs in the highest category of complexity report medians of 11 to 15 annual physician visits and almost 50% report at least one unmet medical service need.(3) Hospitalizations and emergency department (ED) visits can also result in changes to their care plans and additional care coordination needs.

The association between COC and better outcomes has been validated in multiple pediatric studies using the Bice-Boxerman COC index (10), which measures the extent to which a patient's visits are concentrated in a single provider or a team of providers. The Bice-Boxerman COC index has multiple advantages over other COC measures. First, the Bice-Boxerman COC index is sensitive to continuity with a small group of providers, whereas other measures assess continuity with a single provider. For example, if a patient sees two providers frequently, the Bice-Boxerman COC index will reflect that the patient has a lower level of continuity than another patient who sees one provider frequently but will detect the continuity experienced with the two providers. Second, the Bice-Boxerman COC index uses administrative data. The Bice-Boxerman COC index is therefore more feasible than parent-reported measures and is not subject to recall bias. In adults, patient-reported measures of continuity are only moderately correlated with administratively derived measures of continuity – skewing toward high visit continuity.(11)

Prior studies in pediatric populations have found that having a higher Bice-Boxerman COC index in a patient- and family-centered medical home is associated with lower ED utilization(4) and greater primary care provider involvement in care coordination activities, such as communication with other providers, as measured by caregiver report.(5) Conversely, lower continuity is associated with higher risk of hospitalizations and ED utilization.(6)

There is also evidence that having a higher Bice-Boxerman COC index is associated with better outcomes for children with chronic conditions. Among children with at least 1 chronic disease, those with higher COC have lower risk of ambulatory-care sensitive hospitalizations than those with lower COC.(7) Among children with type 1 diabetes, those with higher COC have lower risk of hospitalization for diabetic ketoacidosis.(8) In a study of children with medical complexity, our center found that higher COC was significantly associated with lower ED utilization and more frequent receipt of needed care coordination services.(9)

References

1. Bethell CD, Read D, Blumberg SJ, Newacheck PW. What is the prevalence of children with special health care needs? Toward an understanding of variations in findings and methods across three national surveys. Maternal and child health journal. 2008;12(1):1-14.
2. Berry JG, Hall M, Neff J, et al. Children with medical complexity and Medicaid: spending and cost savings. Health affairs (Project Hope). 2014;33(12):2199-2206.
3. Kuo DZ, Cohen E, Agrawal R, Berry JG, Casey PH. A national profile of caregiver challenges among more medically complex children with special health care needs. Archives of pediatrics & adolescent medicine. 2011;165(11):1020-1026.
4. Christakis DA, Wright JA, Koepsell TD, Emerson S, Connell FA. Is greater continuity of care associated with less emergency department utilization? Pediatrics. 1999;103(4 Pt 1):738-742.
5. Christakis DA, Wright JA, Zimmerman FJ, Bassett AL, Connell FA. Continuity of care is associated with well-coordinated care. Ambulatory pediatrics : the official journal of the Ambulatory Pediatric Association. 2003;3(2):82-86.
6. Christakis DA, Mell L, Koepsell TD, Zimmerman FJ, Connell FA. Association of lower continuity of care with greater risk of emergency department use and hospitalization in children. Pediatrics. 2001;107(3):524-529.
7. Tom JO, Tseng C-W, Davis J, Solomon C, Zhou C, Mangione-Smith R. Missed well-child care visits, low continuity of care, and risk of ambulatory care-sensitive hospitalizations in young children. JAMA Pediatrics. 2010; 11:1052-1058.
8. Christakis DA, Feudtner C, Pihoker C, Connell FA. Continuity and quality of care for children with diabetes who are covered by medicaid. Ambulatory pediatrics : the official journal of the Ambulatory Pediatric Association. 2001;1(2):99-103.
9. Arthur KC, Mangione-Smith R, Burkart Q, Parast L, Liu H, Elliott MN, McGlynn EA, Schneider EC. Association between Continuity of Care and Healthcare Outcomes for Children with Medical Complexity. Under review. 2016.
10. Bice TW, Boxerman SB. A quantitative measure of continuity of care. Med Care. 1977 Apr;15(4):347-9.
11. Rodriguez HP, Marshall RE, Rogers WH, Safran DG. Primary care physician visit continuity: a comparison of patient-reported and administratively derived measures. J Gen Intern Med. 2008 Sep;23(9):1499-1502.

Numerator Statement: Number of eligible children(1) who have a Bice-Boxerman COC index >=0.50 in the primary care setting during the measurement year.

1. Eligible children are defined as children who are continuously enrolled for 12 months with no more than a 30-day gap in enrollment. Children with a gap greater than 30 days are excluded because of the potential for them to be enrolled in a different health plan at that time. In such cases, the child's administrative data for the health plan being measured would be incomplete and thus might not reflect the health plan's true performance on the measure. The timeframe of 30 days as the length of the gap was chosen to be consistent with the month-to-month eligibility assessments used by many Medicaid health plans.

Denominator Statement: Children with medical complexity(1) who are 1-17 years old(2) and who have had >= 4 primary care visits(3) during the measurement year.

1. Children with medical complexity are defined as children who are classified by the Pediatric Medical Complexity algorithm, Version 2 (PMCA-V2) as having no chronic illness or non-complex chronic illness.

2. Children must be >=1 year and <=17 years of age on the last day of the measurement year.

3. Research has shown that stability of the COC index increases as the number of visits increases (ie. less subject to significant change as a result of minor variations in care dispersion).(1) We therefore established a minimum of four visits as has been done in previous studies.(1-3)

References

1. Christakis DA, Wright JA, Koepsell TD, Emerson S, Connell FA. Is greater continuity of care associated with less emergency department utilization? Pediatrics. 1999;103(4 Pt 1):738-742.
2. Christakis DA, Mell L, Koepsell TD, Zimmerman FJ, Connell FA. Association of lower continuity of care with greater risk of emergency department use and hospitalization in children. Pediatrics. 2001;107(3):524-529.
3. Tom JO, Tseng C-W, Davis J, Solomon C, Zhou C, Mangione-Smith R. Missed well-child care visits, low continuity of care, and risk of ambulatory care-sensitive hospitalizations in young children. Arch Pediatr Adolesc Med. 2010; 11:1052-1058.

Denominator Exclusions:

Measure Type: Structure
Data Source: Claims (Only)
Level of Analysis: Health Plan

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence

The developer provides the following evidence for this measure:

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

Evidence Summary

- The developer provides a conceptual framework diagram of the role of continuity of care in care coordination/fragmentation and its relationship with long- and short-term health outcomes for children with medical complexity (CMC).
- The developer "conducted a systematic review of the literature assessing links between continuity of care and health and healthcare outcomes." The systematic review identified 7 studies published from 1999-2010. All reviewed studies utilized the Bice and Boxerman continuity of care index (COC).
- Evidence Summary (7 studies - 6 retrospective cohorts, 1 economic model)
    - Evidence Summary Table
    - Higher continuity of primary care associated with:
        - lower emergency department (ED) utilization
        - lower ED expenditures
        - fewer hospitalizations
        - better parent-reported care coordination
        - better care experiences
        - better diabetes outcomes in children with Type 1 diabetes
    - Lower continuity of primary care associated with:
        - higher ED utilization
        - more hospitalizations.

Questions for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*
- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*

Guidance from the Evidence Algorithm: Structure measure based on systematic review (Box 3) → Empirical evidence without grading of evidence submitted (Box 7) → Summarized empirical evidence includes all studies (Box 8) → Evidence: high certainty of benefits (Box 9) → Moderate

The highest eligible rating is MODERATE

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

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

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

- This is a new measure, so widespread implementation data are not available.
- Based on data from 17 states/test sites, the developer reports percentage of eligible children with a Bice-Boxerman COC index of >=0.50):
  - Pooled rate of 66.3%
  - Range of 23.08% to 96%
  - Average state score of 66.8%
- The developer reports of the 10,860,713 individuals enrolled in Medicaid in the 17 states (2008) included in its analyses, 11,438 were identified as met the age criterion (>=1 year old and <=17 years old) and number of primary care visits (at least 4) to be included in the analysis.
- The overall mean Bice-Boxerman COC score was 0.65 on the 0 to 1.0 scale with standard deviation (SD) of 0.35.

Disparities

- During testing, the developer "assessed performance variation across sociodemographic characteristics using bivariate logistic regression (each characteristic was modeled without adjusting for other covariates)"
- "Children in the youngest age group (0-1 years old) were significantly more likely to pass the measure than children in all other age groups, with the lowest pass rate in adolescents (13 to 17 years old)."

Child age	COC pass rate (%)	p-value
0-1 (reference group)	67.40	
2-5	65.17	<.05
6-12	62.86	<.05
13-17	61.40	<.01

- "Children living in a Metropolitan Statistical Area (MSA) were significantly less likely to pass the measure."

Lives in Metropolitan Statistical Area (MSA)	COC pass rate (%)	p-value
No (reference group)	69.27	
Yes	65.77	<.01

- "Children in the Temporary Assistance for Needy Families (TANF)/low-income category were less likely to pass the measure."

Temporary Assistance for Needy Families (TANF)/Low-income	COC pass rate (%)	p-value
No (reference group)	68.72	
Yes	65.72	<.01

- "No significant differences were observed when examining gender and race/ethnicity."

Child gender	COC pass rate (%)	p-value
Female (reference group)	66.48	
Male	66.14	not significant

Child race/ethnicity	COC pass rate (%)	p-value
Asian/Pacific Islander	69.81	not significant
African American	64.81	not significant
Hispanic	66.20	not significant
White (reference group)	67.07	
Other/multiracial	65.02	not significant

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

Committee pre-evaluation comments

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

- Author provides a review of several studies to support the measure. but i'm curious about direct relationship to patient outcomes. i also am curious about differentiating outcomes with more specifically defining what the 'medical complexities' are.
- It is noted that a small percentage of children with medical complexity account for a large percentage of costs. This structure measure will improve health outcomes and should mention lower costs. It is initially unclear if this is Medicaid only and what happens if the child switches plans or types of plans. It is concerning that neither the quality/quantity/consistency of evidence nor evidence grading is used. The age range in the numerator needs to be consistent with the denominator. Rating: moderate.
- gap exists including those living in metro areas and low income.
- Again it is unclear if this is Medicaid only or if the background evidence applies to Medicaid. It is noted that there are disparities based on age, geographic area, TANF, but not race/ethnicity. Rating: moderate.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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

Data source(s): Claims (Only)

Specifications:

- The level of analysis is health plan.
- Interpretation of score: Better quality = higher score
- The numerator is the "*Number of eligible children who have a Bice-Boxerman COC index >=0.50 in the primary care setting during the measurement year*."
- The denominator is "*Children with medical complexity(1) who are 1-17 years old and who have had >= 4 primary care visits during the measurement year.*
  1. *Children with medical complexity are defined as children who are classified by the Pediatric Medical Complexity algorithm, Version 2 (PMCA-V2) as having no chronic illness or non-complex chronic illness.*
  2. *Children must be >=1 year and <=17 years of age on the last day of the measurement year.*

3. *Research has shown that stability of the COC index increases as the number of visits increases (i.e. less subject to significant change as a result of minor variations in care dispersion). We therefore established a minimum of four visits as has been done in previous studies."*

- There are no exclusions.
- The calculation algorithm is provided.
- The developer provided basic guidelines for the sampling; a minimum of 25 eligible individuals is recommended.
- The measure is not risk adjusted or stratified.
- A data dictionary and list of ICD-9 and ICD-10 codes are provided.

Questions for the Committee:

- *Is the calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing, Testing attachment

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

SUMMARY OF TESTING

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

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

Method(s) of reliability testing

- The developer calculated the degree to which measures scores are able to reliably distinguish between performance levels of state Medicaid agencies.
- The developer examined inter-unit reliability (IUR) for the measure using intra-class correlations (ICCs) computed from individual scores. The developer used a mixed effects logistic regression model with the individual measure score as the outcome (0/1) and random effects for state and obtained the ICC calculated from this model.
- The developer then used ICC to calculate the reliability of the measure based on the median number of eligible individuals per state using the Spearman-Brown formula.

Results of reliability testing

Table 2a2.3.a Reliability results

	Expected Reliability[1]		
	0.7	0.8	0.9
Number of Eligible Individuals per Health Plan Needed to Achieve Expected Reliability	6	11	24

[1] using Spearman-Brown prediction formula

The developer states, **"**These results demonstrate that with a reliability of 0.98, this measure if highly reliable and can be used to distinguish between states in terms of performance. In addition, only 6 eligible individuals per state Medicaid agency (or commercial health plan) are necessary to result in adequate reliability (0.7) of the measure. High reliability of 0.9 can be achieved with 24 eligible individuals per state. This should be feasible, at least at the state Medicaid agency level, since we observed a median of 246 eligible individuals per state Medicaid agency in our analysis (range of 64-

3,730). In summary, these results demonstrate that we expect this measure to be able to reliably distinguish between state Medicaid agencies (or health plans) in terms of performance."

Questions for the Committee:

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

Guidance from the Reliability Algorithm: Precise specifications (Box 1)→ Empirical reliability testing conducting (Box 2)→ Computed performance measure scores used (Box 4)→ Appropriate method for assessing variability (Box 5)→ High certainty (Box 6a)→ High

The highest eligible rating is HIGH

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

2b. Validity

2b1. Validity: Specifications

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

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- o *Are the specifications consistent with the evidence?*

2b2. Validity testing

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

SUMMARY OF TESTING

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

Method of validity testing of the measure score:

- ☒ **Face validity**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- Empirical validity testing
  - o Analysis of the relationship between the COC quality measure score and the occurrence of at least one ED visit in the measurement year at individual level
  - o Mixed-effects logistic regressions that predicted the occurrence of at least one ED visit from the measure score, with random intercepts for states.
  - o Analysis of unadjusted and adjusted models for characteristics associated with both the COC quality measure and the occurrence of at least one ED visit.
- Face validity
  - o For face validity, NQF requires an assessment of whether the measure score has been judged as able to distinguish good from bad quality.
  - o A 9-member multi-stakeholder Delphi panel was assembled. The panelists read the literature reviews written by the project staff and scored the each proposed quality measure based on validity from a scale from 1 (low) to 9 (high).

- The developer reports that among the rating factors considered was "whether [the rater] would consider providers who adhere more consistently to the quality measure to be providing higher quality care."

Validity testing results:

- The COC quality measure was significantly associated with decreased odds of having one or more ED visits ($p<0.05$) Table 2b2.3.a shows the odds ratio (OR) and 95% confidence limits of an ED visit for children with a Bice-Boxerman COC index of 0.5 or greater (COC quality measure pass) compared to children with a Bice-Boxerman COC index below 0.5 (COC quality measure fail).

Table 2b2.3.a Empirical validity results

	Odds ratio (OR)	Lower Limit of 95% confidence interval (LCL)	Upper Limit of 95% confidence interval (UCL)	p-value
Unadjusted				
Bice-Boxerman COC index of 0.5 or greater (COC quality measure pass)	0.90	0.83	0.98	0.0110
Adjusted for SDS				
Bice-Boxerman COC index of 0.5 or greater (COC quality measure pass)	0.88	0.81	0.95	0.0021

- The developer concludes the measure demonstrates face validity (Table 2b2.3.b).

Table 2b2.3.b Delphi Panel results (aggregate result provided, not on the individual components for which face validity was assessed)

	Median score (Scale 1-9)	Mean absolute deviation from median	Agreement status*
Face Validity	7.0	0.7	Indeterminate
Feasibility	7.0	0.6	Agreement

*This is a statistical assessment of whether panelists agreed, disagreed, or if status was indeterminate.

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer lists as N/A.

Question for the Committee:

- *Is the lack of exclusions consistent with the evidence?*

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

- The developer notes it does "not recommend risk adjustment of the COC quality measure for SDS when it is used to compare health plan performance….We fit the adjusted validation model to remove potential

confounders of the relationship between the COC quality measure and the validation metric in our particular test dataset."

Question for the Committee:

- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this structural measure for SDS factors?*

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

- 17 state Medicaid agencies were reviewed using a chi-square omnibus test.
- The developer was able to "detect statistically and clinically meaningful differences in state Medicaid agency performance," as well as "statistically and clinically meaningful differences in COC quality measure scores across demographic subgroups for the following characteristics: child age, residence in an MSA, and receipt of TANF."

Table 2b5.2.a Differences in COC Quality Measure Scores by State Medicaid Agencies

	Denominator	Numerator	COC Quality Measure Pass Rate (%)	P-value for omnibus test	Difference from overall mean of others	P-value for difference from overall mean of others
17 States pooled	11438	7583	66.30	<.0001	--	--
AZ	478	333	69.67	--	3.52	0.1118
CA	699	513	73.39	--	7.55	<.0001
IL	1459	1113	76.29	--	11.44	<.0001
IN	858	785	91.49	--	27.24	<.0001
KS	154	89	57.79	--	-8.62	0.0254
KY	242	77	31.82	--	-35.22	<.0001
LA	84	65	77.38	--	11.15	0.0334
MN	64	51	79.69	--	13.43	0.0263
NC	450	432	96.00	--	30.92	<.0001
NE	84	80	95.24	--	29.15	<.0001
NJ	171	135	78.95	--	12.84	0.0005
NM	246	90	36.59	--	-30.36	<.0001
NY	1950	1272	65.23	--	-1.29	0.2715
TN	469	256	54.58	--	-12.21	<.0001
TX	3730	2136	57.27	--	-13.40	<.0001
VA	183	129	70.49	--	4.26	0.2269
WI	117	27	23.08	--	-43.65	<.0001

Table 2b5.2.b Differences in COC Quality Measure Scores by Sociodemographic Characteristics

	COC Quality Measure Pass Rate (%)	p-value
Child age		
0-1 (reference group)	67.40	
2-5	65.17	0.0264

6-12	62.86	0.0246
13-17	61.40	0.0065
Child gender		
Female (reference group)	66.48	
Male	66.14	0.6974
Child race/ethnicity		
Asian/Pacific Islander	69.81	0.2578
African American	64.81	0.1227
Hispanic	66.20	0.4117
White (reference group)	67.07	
Other/multiracial	65.02	0.2609
Lives in MSA		
No (reference group)	69.27	
Yes	65.77	0.0046
TANF/Low Income		
No (reference group)	68.72	
Yes	65.72	0.0076

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- N/A

2b7. Missing Data

- Individuals with inadequate claims that PMCA were not able to assess for eligibility were not included in analysis.
  - 6.6% of the source population (n=715,609) had inadequate claims and were not included in analysis.
- The developer states it does not expect missing data to effect reliability and validity robustness. The developer acknowledges missing data can introduce bias, but further notes it does not expect an impact any different from that which occurs for measures that rely on claims data.

Guidance from the Validity Algorithm: Threats to validity assessed? (Box 2)→ Empirical validity tested using measures as specified? (Box 3)→ Validity testing computed performance measure scores for each measured entity (Box 6)→ Appropriate method used? (Box 7)→ Moderate certainty (Box 8b)→ High

The highest eligible rating is HIGH.

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

Committee pre-evaluation comments

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

- It is noted that there are no exclusions but clarification is needed regarding continuous enrollment (e.g. switch plans). It is noted that the specifications are consistent with the evidence.
- developer provides information how data reliably distinguishes performance between states of patients on medicaid plans across the country.
- It is also noted that the reliability testing was performed with the data source and level of analysis required. Rating: moderate.
- Face validity and empirical validity. high validity demonstrated using algorithm.
- It is noted that measure score only is used. Although the developer notes demonstrated validity, face validity was indeterminate.

- 6% had inadequate claims and were not included in the analysis. developer does not think that this would change validity results.
- It is noted under 2b3 that there are no exclusions but it is uncertain what occurs if there is continuous enrollment but change in plans. This is also a concern for risk adjustment under 2b4. Under 2b5 it is reassuring that statistically and clinically meaningful differences were detected. For 2b6 it is unclear why there is no comparability of data sources. Under 2b7, it is agreed that not including claims that were incomplete will not affect reliability and validity. Rating: moderate.

Criterion 3. Feasibility

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

- All data elements are defined in fields in electronic claims.
- The data elements will be "coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)"
- There are no associated fees, licensing or other requirements to use any aspects of the measure as specified.
- The measure requires SAS codes, available from the developer, to compute.

Questions for the Committee:

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

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

Committee pre-evaluation comments

Criteria 3: Feasibility

- Electronic claims. i wonder about how labor intensive it is to extract data.
- The acronym SAS needs to be defined. It is agreed that the required data elements are routinely generated, available in electronic format, and ready to be operational. Rating: moderate.

Criterion 4: Usability and Use

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

Accountability program details

- The measure has been developed as part of the Pediatric Quality Measurement Program, funded by AHRQ and CMS, using CHIPRA monies.
- The Minnesota and Washington State Medicaid agencies will work with the National Academy for State Health Policy and the Medicaid Medical Directors Learning Network (MMDLN) to encourage state Medicaid agencies to implement the measure.

Improvement results

- The measure has not yet been implemented.

Unexpected findings (positive or negative) during implementation

- None observed by the developer

Potential harms

- None observed by the developer

Vetting of the measure

- The measure has not been vetted.

Feedback:

- No feedback was obtained.

***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?*
- *How has the measure been vetted in real-world settings by those being measure or others?*

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

Committee pre-evaluation comments

Criteria 4: Usability and Use

- "this is for planned use in accountability program.
  - -has not been vetted.
  - -how is 'small group of providers' defined?
  - -for 'urgent care visits,' what if it's not possible to see the primary care provider, but that person is available by phone or 'curbside' to collaborate? "
- It is concerning that this is not publicly reported. It is agreed this measure would further the goal of high quality health care by improving outcomes. This measure was not vetted. There is no section 5 so will report here that this measure is not eligible for endorsement and designation.

Criterion 5: Related and Competing Measures

Related or competing measures

- N/A

Harmonization

- N/A

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

RATIONALE IF NOT ELIGIBLE:

The measure has not been vetted by those being measured or other users.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Continuity of Primary Care for Children with Medical Complexity
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: Click here to enter a date

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

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

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Health outcome: Click here to name the health outcome

☐Patient-reported outcome (PRO): Click here to name the PRO

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

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

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

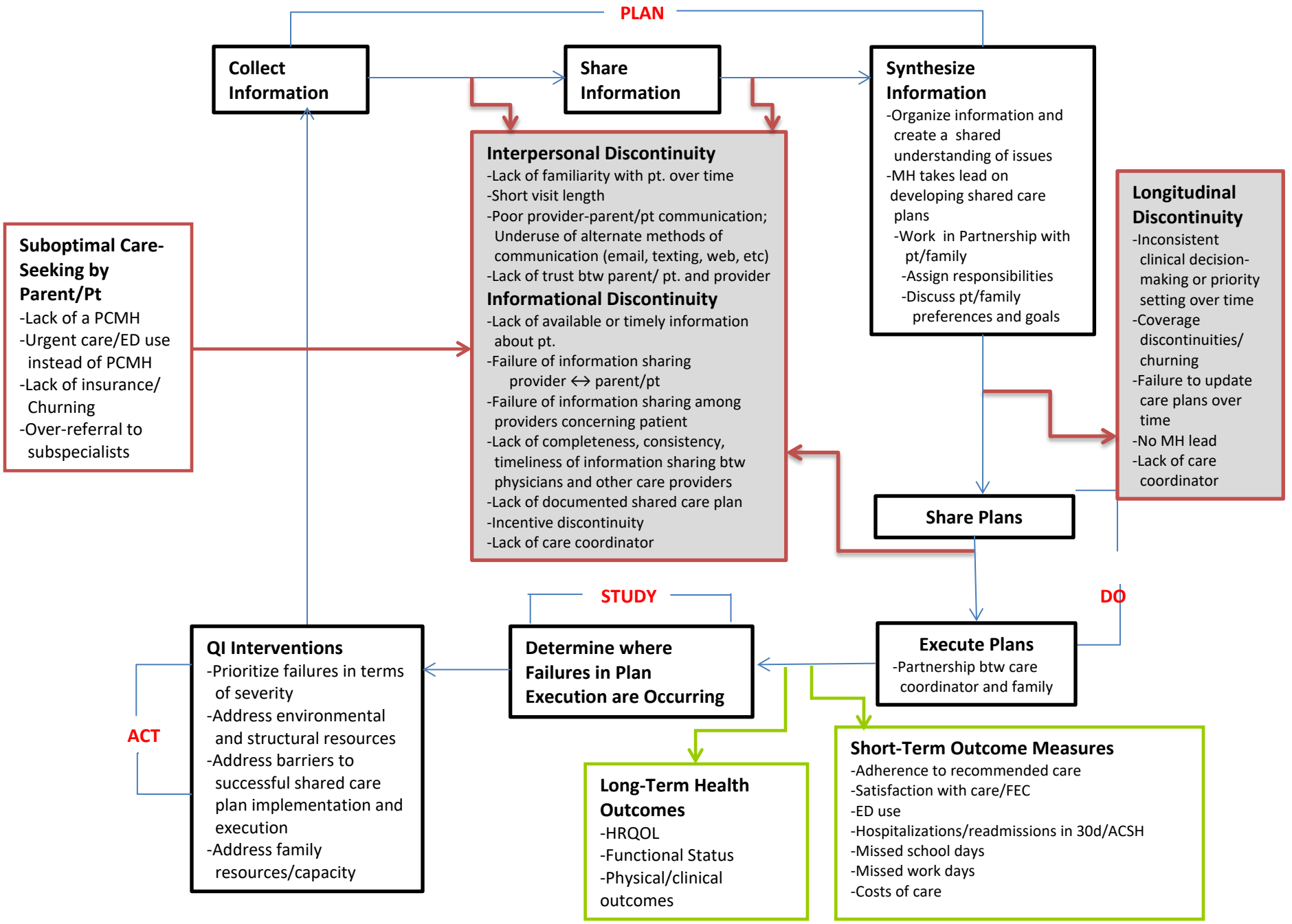
☒ Structure: Continuity of primary care for children with medical complexity

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

The conceptual framework below diagrams the role of continuity of care in care coordination/fragmentation and its relationship with long- and short-term health outcomes for children with medical complexity (CMC). Continuity of primary care is considered foundational to all processes in the plan, do, study, act cycle; the gray highlighted boxes represent "voltage drops" from the main cycle when there is discontinuity. In the validation analyses using administrative claims data, we focus on two outcomes: emergency department visits and ambulatory care-sensitive hospitalizations.

Conceptual Framework for Care Coordination/Fragmentation in the Context of the PCMH for Children with Complex Needs



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

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence:	

• Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

_____

1a.4 OTHER SOURCE OF EVIDENCE

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

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

We identified seven studies examining the association between continuity of primary care and child health and healthcare outcomes and one study examining patient characteristics associated with valuing continuity of care in our systematic review. Details about each study are provided below, followed by a table summarizing the study designs and findings for the 9 studies examining the association with health and healthcare outcomes. In brief, these studies found that higher continuity of primary care was associated with better outcomes, including lower emergency department (ED) utilization, lower ED expenditures, fewer hospitalizations, better parent-reported care coordination, better care experiences, and better diabetes outcomes in children with Type 1 diabetes. Conversely, lower continuity of primary care was associated with higher ED utilization and more hospitalizations. The quality of the evidence reported in this synthesis is level 2 (cohort studies).

All of the studies used the Bice and Boxerman continuity of care (COC) index. In the studies examining ED utilization and hospitalizations, the COC index was calculated for each patient and modeled as a categorical variable with categories ranging from low to high on a 0-1 scale (See Table 1). Because of variation in the length of each child's enrollment

and the time at which the child used the ED or was hospitalized, these studies used Cox proportional hazards regression models. Note that our proposed quality measure on Continuity of Primary Care for Children with Medical Complexity establishes a minimum level of 0.5 on the 0-1 scale.

Table 1. Levels of COC index used in various studies

	High	Medium-high	Medium	Low
Christakis et al.[1]	0.41 - 1.0	-	0.19 - 0.40	0 - 0.18
McBurney et al.[2]	0.48 - 1.0	-	0.26 - 0.47	0 - 0.18
Tom et al.[3]	0.75 - 1.0	0.51 - 0.74	0.26 - 0.50	0 - 0.25

COC and ED Utilization

We identified one study examining the association between lower continuity and ED utilization and one study examining the association between higher continuity and ED utilization. The first, which was a retrospective cohort study by Christakis et al., found that lower continuity is associated with higher ED utilization.[1] The study used claims data from pediatric patients enrolled at a large health maintenance organization in one state (n=46,097), including children with and without chronic conditions. Children in the medium COC category were significantly more likely to visit the ED than children in the high COC category (HR 1.28 [1.20-1.36]), while children in the lowest COC category were even more likely to visit the ED (HR 1.58 [1.49-1.66]) in adjusted analyses. When examining a subset of children with Medicaid insurance, children in the low category were significantly more likely to visit the ED than children in the high continuity category (HR 1.40 [1.02-1.92]).

A second retrospective cohort study by Christakis et al. found that higher continuity was associated with lower ED utilization in a Medicaid managed care pediatric population (n=785).[4] The setting was an outpatient resident teaching clinic where residents were supervised by attending physicians. Lower ED use was associated with medium attending

physician continuity (HR 0.70 [0.53-0.93]) and high attending physician continuity (HR 0.65 [0.50-0.80]). Attending physician continuity appeared to have a dose-response relationship, with high continuity being more strongly associated with lower ED use than medium continuity.

COC and Economic Modeling of Costs of ED Care

McBurney et al. created an economic model for the ED costs avoided when continuity increased.[2] The authors created two hypothetical pediatric medical homes of 2,000 patients each with ED use of 0.68 visits per child per year based on utilization data from three studies examining pediatric ED utilization.[5] [6,7]. In the economic model, the first practice had an average of 40% continuity with patients distributed across low, medium and high categories of the COC index, while the second practice had an average of 50% continuity with differing numbers of patients distributed across the same categories. The 10 percentage point increase in continuity resulted in a decline of expected ED visits from 1,362 to 1,290 with savings of $19,905 per 2,000 patients. When repeating the model using the hazard ratios published by Christakis et al. in a 1999 study of COC and ED utilization using the Bice and Boxerman COC index,[1] 85 ED visits were avoided and $23,519 were saved in a one-year period for 2,000 patients.

COC and Hospitalization

Two retrospective cohort studies have found an association between low continuity and greater risk of hospitalization.[1,3] Christakis et al. assessed risk of hospitalization in pediatric patients—including children with chronic conditions—enrolled at a large health maintenance organization (n=46,097).[1] Children in the medium COC category were more likely to be hospitalized than children in the high COC category (HR 1.22 [1.09-1.38]), while children in the lowest COC category were even more likely to be hospitalized (HR 1.54 [1.33-1.75]) in adjusted analysis. When examining a subset of children with Medicaid

insurance, children in the low continuity category were significantly more likely to be hospitalized (HR 1.61 [1.01-3.03]) in adjusted analysis.

Tom et al. studied the association between continuity and ambulatory care-sensitive hospitalization (ACSH) in a retrospective cohort study in children under 3.5 years old (n=36,944) enrolled in a health plan in which 24% of the children had ≥1 chronic diseases.[3] ACSH are hospitalizations for conditions that are more amenable to primary care interventions[8] and can be further classified as chronic, e.g. an asthma exacerbation, or acute, e.g. dehydration. The authors defined children with chronic diseases as having at least one claim in the study period with a diagnosis included in a validated list of ICD-9-CM chronic disease codes for children.[9] Among children with ≥1 chronic diseases, the risk of ACSH was twice as high for children in the lowest COC category compared to those who were in the highest category (HR 2.4 [1.7-3.5]).

COC and Type 1 Diabetes Healthcare Outcomes

A third retrospective cohort study by Christakis et al. examined continuity of care in a Medicaid-insured pediatric population with Type 1 diabetes mellitus (n=252).[10] Children with medium COC and high COC were less likely to be hospitalized for diabetic ketoacidosis and less likely to have diabetic ketoacidosis as an outpatient. Children with high COC were also more likely to have visited an ophthalmologist.

COC and Parent Reports of Coordination of Care

A retrospective cohort study by Christakis et al. found that higher continuity was associated with better parent-reported coordination of care (n=759). Parent-reported experience with care coordination with the child's provider was measured using five items of the Components of Primary Care Index (CPCI). The CPCI included items about the provider's knowledge about the patient's care in other clinics, communication with other providers, knowledge about the results of visits to other doctors, follow-up communication with the family, and preference to have one provider coordinate care.[11] Greater continuity

was associated with statistically significantly higher CPCI ratings for the care coordination domain as a whole and for all five items individually in adjusted analyses.

Continuity and Parent-Reported Satisfaction

Christakis et al. found that greater continuity was associated with improved family experience in a retrospective cohort study (n=759).[12] The study used six provider-specific items from the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) survey and two overall ratings (of the clinic and the provider) as outcome variables. Greater continuity of care was associated with significantly higher ratings for five of the six provider-specific items as well as a higher overall clinic rating and a higher overall provider rating after adjusting for the number of visits, the length of time enrolled at the clinic, race/ethnicity, age of child, gender of child and income.

Characteristics Associated with Placing Value on COC

Finally, in a study of adult and pediatric patients at 84 practices (n=4,454), Nutting et al. found that several patient characteristics were associated with placing higher value on continuity, including extremes of age (such as parents of children under age 6), female sex, less education, Medicare and Medicaid insurance, number of chronic conditions and medications, number of visits to the practice and worse self-reported health status ($p<.001$ for all characteristics except sex, where $p<.015$).[13] Given that children with medical complexity by definition have more than one chronic condition and likely have more medications, more visits and worse health status, these findings provide further evidence of the potential importance of continuity of primary care for the Medicaid-insured population targeted by this quality measure.

Evidence Summary Table: Outcomes Observed with Higher Continuity

Source and Study Design	Population	Measure of continuity	ED utilization	Hospitalizations	Chronic Ambulatory care-sensitive hospitalizations	Acute Ambulatory care-sensitive hospitalizations	Cost of ED care	Care Coordination	Satisfaction
Christakis 2001[1] Retrospective cohort	Pediatric patients in a health management organization in Washington state (n=46,097)	COC index	↓	↓					
Christakis 1999[4] Retrospective cohort	Medicaid managed care pediatric patients at outpatient resident teaching clinic in Washington state (n=785)	COC index	↓						
McBurney 2004[2] Economic model	Model based on results from previously published studies on pediatric patients	COC index (economic model)					↓		
Tom 2010[3] Retrospective cohort	Children age 3.5 or younger enrolled in Hawaii's largest health plan from 1999 to 2006 (n=36,944)	COC index		↓	↓	↓			
Christakis 2001[10] Retrospective cohort	Children with type 1 diabetes ensured by Medicaid in Washington state (n=252)	COC index		↓*					

Source and Study Design	Population	Measure of continuity	ED utilization	Hospitalizations	Chronic Ambulatory care-sensitive hospitalizations	cute Ambulatory care-sensitive hospitalizations	Cost of ED care	Care Coordination	Satisfaction
Christakis 2003[11] Retrospective cohort	Pediatric patients from a primary care clinic and a resident teaching clinic in Washington state (n=759)	COC index						↑	
Christakis 2002[12] Retrospective cohort	Pediatric patients from a primary care clinic and a resident teaching clinic in Washington state (n=759)	COC index							↑

ED = Emergency department
↑ = statistically significant increase in outcome measure
↓ = statistically significant decrease in outcome measure

*Christakis et al. examined hospitalizations for diabetic ketoacidosis.

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

We began by developing a conceptual framework for care coordination/fragmentation for children with medical complexity (see 1a.12, above). The framework shows the many ways in which longitudinal discontinuity, interpersonal discontinuity and informational discontinuity contribute to fragmentation and poor care coordination. Longitudinal continuity is defined as receiving ongoing care from one provider or one team of providers in the same location, whereas interpersonal continuity is a type of longitudinal continuity that refers to having an ongoing personal relationship between a patient and a care provider with the additional characteristic of personal trust and responsibility.[14] Informational continuity refers to the availability of information about a patient's health care encounters to all the providers involved in a patient's care even at different locations.[14] The framework also illuminates how continuity of care relates to both short- and long-term outcomes, such as emergency department utilization and satisfaction with care.

Based on this framework, we conducted a systematic review of the literature assessing links between continuity of care and health and healthcare outcomes. We then drafted quality measures that were supported by the identified evidence.

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

1. Christakis DA, Mell L, Koepsell TD, Zimmerman FJ, Connell FA. Association of lower continuity of care with greater risk of emergency department use and hospitalization in children. *Pediatrics.* 2001;107(3):524-529.
2. McBurney PG, Simpson KN, Darden PM. Potential cost savings of decreased emergency department visits through increased continuity in a pediatric medical home. *Ambulatory pediatrics : the official journal of the Ambulatory Pediatric Association.* 2004;4(3):204-208.
3. Tom JO, Tseng C-W, Davis J, Solomon C, Zhou C, Mangione-Smith R. Missed well-child care visits, low continuity of care, and risk of ambulatory care-sensitive hospitalizations in young children. 2010; 11:1052-1058.
4. Christakis DA, Wright JA, Koepsell TD, Emerson S, Connell FA. Is greater continuity of care associated with less emergency department utilization? *Pediatrics.* 1999;103(4 Pt 1):738-742.
5. Gill JM, Mainous AG, 3rd, Nsereko M. The effect of continuity of care on emergency department use. *Archives of family medicine.* 2000;9(4):333-338.
6. DeAngelis C, Fosarelli P, Duggan AK. Use of the emergency department by children enrolled in a primary care clinic. *Pediatric emergency care.* 1985;1(2):61-65.
7. Chessare JB. Utilization of emergency services among patients of a pediatric group practice. *Pediatric emergency care.* 1986;2(4):227-230.
8. IOM (Institute of Medicine). *Access to Health Care in America.* Washington, DC: National Academy Press; 1993.
9. Kuhlthau KA, Beal AC, Ferris TG, Perrin JM. Comparing a diagnosis list with a survey method to identify children with chronic conditions in an urban health center. *Ambulatory pediatrics : the official journal of the Ambulatory Pediatric Association.* 2002;2(1):58-62.
10. Christakis DA, Feudtner C, Pihoker C, Connell FA. Continuity and quality of care for children with diabetes who are covered by medicaid. *Ambulatory pediatrics : the official journal of the Ambulatory Pediatric Association.* 2001;1(2):99-103.

11. Christakis DA, Wright JA, Zimmerman FJ, Bassett AL, Connell FA. Continuity of care is associated with well-coordinated care. *Ambulatory pediatrics : the official journal of the Ambulatory Pediatric Association.* 2003;3(2):82-86.
12. Christakis DA, Wright JA, Zimmerman FJ, Bassett AL, Connell FA. Continuity of care is associated with high-quality careby parental report. *Pediatrics.* 2002;109(4):e54.
13. Nutting PA, Goodwin MA, Flocke SA, Zyzanski SJ, Stange KC. Continuity of primary care: to whom does it matter and when? *Annals of family medicine.* 2003;1(3):149-155.
14. Saultz JW. Defining and measuring interpersonal continuity of care. *Annals of family medicine.* 2003;1(3):134-143.

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

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

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

NQF_COC_evidence_attachment_FINAL_12-5-16.docx

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

Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

1b. Performance Gap

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

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

1b.1. **Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*
<u>IF a PRO-PM</u> (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)
<u>IF a COMPOSITE</u> (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.

Increasing numbers of children in the United States are living with medical complexity.(1) An analysis of the 2005-2006 National Survey of Children with Special Health Care Needs found that children in the highest category of complexity represent 0.4% of all children in the United States. In a 2014 study in 12 states, children with medical complexity (CMC) represented only 5.8% of children covered by Medicaid but accounted for 34% of all healthcare expenditures.(2) Given the challenges of coordinating care for these children, a continuous relationship with a single primary care provider or small number of providers with in-depth knowledge of their medical and social needs is essential.

CMC stand to benefit from continuity of care because repeated contact could provide more opportunities to discuss their needs and receive help coordinating the many providers and services involved in their care outside of the primary care setting. Nationally, families of children with special needs in the highest category of complexity report medians of 11 to 15 annual physician visits and almost 50% report at least one unmet medical service need.(3) Hospitalizations and emergency department (ED) visits can also result in changes to their care plans and additional care coordination needs.

Prior studies in pediatric populations have found that higher levels of COC in a patient- and family-centered medical home are associated with lower ED utilization(4) and greater primary care provider involvement in care coordination activities, such as communication with other providers, as measured by caregiver report.(5) Conversely, lower continuity is associated with higher risk of hospitalizations and ED utilization.(6)

There is also evidence that COC is associated with better outcomes for children with chronic conditions. Among children with at least 1 chronic disease, those with higher COC have lower risk of ambulatory-care sensitive hospitalizations than those with lower COC.(7) Among children with type 1 diabetes, those with higher COC have lower risk of hospitalization for diabetic ketoacidosis.(8) In a study of children with medical complexity, our center found that higher COC was significantly associated with lower ED utilization and more frequent receipt of needed care coordination services.(9)

References

1. Bethell CD, Read D, Blumberg SJ, Newacheck PW. What is the prevalence of children with special health care needs? Toward an understanding of variations in findings and methods across three national surveys. Maternal and child health journal. 2008;12(1):1-14.
2. Berry JG, Hall M, Neff J, et al. Children with medical complexity and Medicaid: spending and cost savings. Health affairs (Project Hope). 2014;33(12):2199-2206.
3. Kuo DZ, Cohen E, Agrawal R, Berry JG, Casey PH. A national profile of caregiver challenges among more medically complex children with special health care needs. Archives of pediatrics & adolescent medicine. 2011;165(11):1020-1026.
4. Christakis DA, Wright JA, Koepsell TD, Emerson S, Connell FA. Is greater continuity of care associated with less emergency department utilization? Pediatrics. 1999;103(4 Pt 1):738-742.
5. Christakis DA, Wright JA, Zimmerman FJ, Bassett AL, Connell FA. Continuity of care is associated with well-coordinated care. Ambulatory pediatrics : the official journal of the Ambulatory Pediatric Association. 2003;3(2):82-86.
6. Christakis DA, Mell L, Koepsell TD, Zimmerman FJ, Connell FA. Association of lower continuity of care with greater risk of emergency department use and hospitalization in children. Pediatrics. 2001;107(3):524-529.
7. Tom JO, Tseng C-W, Davis J, Solomon C, Zhou C, Mangione-Smith R. Missed well-child care visits, low continuity of care, and risk of ambulatory care-sensitive hospitalizations in young children. JAMA Pediatrics. 2010; 11:1052-1058.
8. Christakis DA, Feudtner C, Pihoker C, Connell FA. Continuity and quality of care for children with diabetes who are covered by medicaid. Ambulatory pediatrics : the official journal of the Ambulatory Pediatric Association. 2001;1(2):99-103.
9. Arthur KC, Mangione-Smith R, Burkart Q, Parast L, Liu H, Elliott MN, McGlynn EA, Schneider EC. Association between Continuity of Care and Healthcare Outcomes for Children with Medical Complexity. Under review. 2016.

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

We obtained Medicaid Analytic eXtract (MAX) data for 17 state Medicaid agencies for 2005-2008. These 17 state Medicaid agencies were selected based on the quality of their Medicaid Analytic eXtract (MAX) data for research purposes.(1) The 17 states included Arizona, California, Illinois, Indiana, Kansas, Kentucky, Louisiana, Minnesota, Nebraska, New Jersey, New Mexico, New York, North Carolina, Tennessee, Texas, Virginia, and Wisconsin.

The Bice-Boxerman COC index measurement year was from Jan. 1, 2008 to Dec. 31, 2008. Data from Jan. 1, 2005 to Dec. 31, 2008 were used to run the Pediatric Medical Complexity Algorithm, Version 2 (PMCA-V2) to identify children with complex chronic conditions (1 of the 4 eligibility criteria for the measure).(2) Data from Jan. 1, 2008 to Dec. 31, 2008 were used to evaluate the remaining 3 eligibility criteria (child age as of December 31 of the measurement year, minimum number of primary care visits, enrollment gaps) and to calculate the Bice-Boxerman COC score.

A total of 10,860,713 individuals in the 17 states used in our analyses were enrolled in Medicaid for the entire measurement year (2008) with no more than a 30-day gap. Of these, 1,095,068 were identified as having complex chronic conditions using PMCA. Of those, 11,438 met the age criterion of being >= 1 year old and < = 17 years old as of December 31, 2008, and had at least 4 primary care visits in 2008 and were included in this analysis.

The overall mean Bice-Boxerman COC score was 0.65 on the 0 to 1.0 scale with standard deviation (SD) of 0.35.

The average state score for the COC quality measure (percentage of eligible children with a Bice-Boxerman COC index of >=0.50) was 66.8%. The minimum state-level score was 23.1% and the maximum state-level score was 96%.

References

1. Byrd & Dodd. 2013. Assessing the usability of MAX 2008 encounter data for comprehensive managed care. Medicare & Medicaid Research Review. 2008 Medicaid Managed Care Enrollment Report, Centers for Medicare and Medicaid Services, U.S. Department of Health and Human Services as of 06/30/2016; available at http://kff.org/medicaid/state-indicator/total-medicaid-mc-enrollment

2. Simon TD, Cawthon ML, Popalisky J, Mangione-Smith R. Development and Validation of the Pediatric Medical Complexity Algorithm (PMCA) Version 2.0. Hospital Pediatrics. In Press.

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

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*
We obtained Medicaid Analytic eXtract (MAX) data for 17 state Medicaid agencies for 2005-2008, as described above. The same 11,438 child enrollees included in the analysis of performance scores by state were included in this analysis examining disparities.

Demographic characteristics of sample

Child age, N(%)
0-1: 7200 (63)*
2-5: 3164 (28)
6-12: 587 (5)
13-17: 487 (4)

Child gender, N (%)
Female: 5352 (47)
Male: 6086 (53)

Child race/ethnicity, N %
Asian/Pacific Islander: 424 (4)
African American: 1563 (14)
Hispanic: 5435 (48)
White: 3167 (28)
Other/multiracial: 849 (7)

Lives in Metropolitan Statistical Area (MSA), N (%)
No: 1731 (15)
Yes: 9707 (85)

Temporary Assistance for Needy Families(TANF)/Low-income,N (%)
No: 1731 (15)
Yes: 9707 (85)

*168 (1.5%) of the sample were 11 months old on December 31, 2008 and had no gaps in enrollment, so were retained in the analysis.

We assessed performance variation across sociodemographic characteristics using bivariate logistic regression (each characteristic was modeled without adjusting for other covariates); our findings are presented below. Children in the youngest age group (0-1 years old) were significantly more likely to pass the measure than children in all other age groups, with the lowest pass rate in adolescents (13 to 17 years old). Children living in a Metropolitan Statistical Area (MSA) were significantly less likely to pass the measure. Children in the Temporary Assistance for Needy Families (TANF)/low-income category were less likely to pass the measure. No significant differences were observed when examining gender and race/ethnicity.

Differences in COC Measure Scores by Sociodemographic Characteristics

The COC pass rate (%) is provided for each category of the sociodemographic characteristics.

Child age
0-1 (reference group): 67.40
2-5: 65.17*
6-12: 62.86*
13-17: 61.40**

Child gender
Female (reference group): 66.48
Male: 66.14 NS

Child race/ethnicity
Asian/Pacific Islander: 69.81 NS
African American: 64.81 NS
Hispanic: 66.20 NS
White (reference group): 67.07
Other/multiracial: 65.02 NS

Lives in Metropolitan Statistical Area (MSA)
No (reference group): 69.27
Yes: 65.77**

Temporary Assistance for Needy Families (TANF)/Low-income
No (reference group): 68.72
Yes: 65.72**

*p<.05
**p<.01
NS not significant

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

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

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

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://www.seattlechildrens.org/research/child-health-behavior-and-development/mangione-smith-lab/measurement-tools/

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

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

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

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

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Number of eligible children who have a Bice-Boxerman COC index >=0.50 in the primary care setting during the measurement year.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Administrative claims data, i.e. CPT codes and ICD-9 (or ICD-10) codes, for all primary care utilization – including both preventive and acute care visits – are needed during the 12-month measurement period to calculate the Bice-Boxerman COC index. The National Provider Identifier (NPI) code is also needed for each primary care visit that occurred during the measurement period.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Children with medical complexity who are 1-17 years old and who have had >= 4 primary care visits during the measurement year

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The details for denominator identification using the PMCA-V2 are provided at http://www.seattlechildrens.org/research/child-health-behavior-and-development/mangione-smith-lab/measurement-tools/, including the ICD-9 codes used for determining PMCA-V2 categorization.

The ICD-9/ICD-10 combined PMCA SAS programming will be available at this website in March of 2017. The draft version is attached as an Appendix to this submission.

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Children who do not meet the criteria outlined in the denominator statement will be excluded, ie. children who:

1. Are >= 1 year and <=17 years of age on the last day (December 31) of the measurement year.
2. Are classified by the Pediatric Medical Complexity Algorithm, Version 2 (PMCA-V2) as having no chronic illness or non-complex chronic illness
3. Have <4 primary care visits in the Bice-Boxerman COC measurement year
4. Are not enrolled in the health plan for 12 months of the Bice-Boxerman COC measurement year and/or have more than a 30-day gap in enrollment in the Bice-Boxerman COC measurement year.

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
1: Dates of birth from administrative claims data are needed to calculate age on December 31 of the measurement year.
2: 1 to 3 years of administrative claims data (including the measurement year) are needed to calculate the PMCA-V2 category for each child.
3: At least 1 year of data with no more than a 30-day gap in enrollment is needed to calculate the Bice-Boxerman COC index for each child.
4: Dates of enrollment are needed to determine whether the child had any gaps in enrollment in the Bice-Boxerman COC index measurement year.

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*
N/A, no stratification is recommended.

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
No risk adjustment or risk stratification
If other:

S.12. **Type of score:**
Rate/proportion
If other:

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

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
To produce scores for the Continuity of Primary Care for Children with Medical Complexity quality measure, the following steps should be taken in this order:

1. Identify child enrollees age >=1 and <=17 on December 31 of the measurement year.
2. Retain those who were continuously enrolled for the 12 months of the measurement year with no more than a 30-day gap in enrollment.
3. Run the PMCA-V2 algorithm and retain only those classified as having complex chronic disease using the SAS programming code available at http://www.seattlechildrens.org/research/child-health-behavior-and-development/mangione-smith-lab/measurement-tools/.
4. Retain those with >=4 primary care visits during the measurement year. The denominator population has now been determined.

5. Calculate the Bice-Boxerman COC index score for eligible child enrollees in the denominator population using the SAS code available at http://www.seattlechildrens.org/research/child-health-behavior-and-development/mangione-smith-lab/measurement-tools/.
6. Calculate the percentage of eligible child enrollees with a Bice-Boxerman COC index >=0.5 by dividing the number of eligible child enrollees with a Bice-Boxerman COC index>=0.5 by the denominator of all eligible children determined by steps 1-4 above.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
To maximize measure reliability, we recommend a minimum sample size of 25 eligible individuals (with sufficient data, i.e. >=4 primary care visits, to calculate the COC measure) calculated per unit of comparison (e.g. Medicaid agency, health plan, etc.). The intraclass correlation for state Medicaid agencies was 0.28 (95% CI 0.16-0.44).

Reliability Results (See also Table 2a2.3.a in the testing attachment.)

To achieve a reliability level of 0.7, 6 eligible children are needed per unit of comparison.
To achieve a reliability level of 0.8, 11 eligible children are needed per unit of comparison.
To achieve a reliability level of 0.9, 24 eligible children are needed per unit of comparison.

Note that in our test data, the median number of patients per state was 246 (minimum 64, maximum 3,730).

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A; measure is not based on a survey or instrument.

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Claims (Only)

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
Denominator: ICD-9 or ICD-10 codes are needed during the 12-month measurement period (January 1 to December 31) to identify children with complex conditions using PMCA-V2.

Numerator: Administrative claims data, ie. CPT codes and ICD-9/ICD-10 codes, for all primary care utilization - including both preventive and acute care visits - are needed during the 12-month measurement period (January 1 to December 31) to calculate the Bice-Boxerman COC index. The National Provider Identifier (NPI) code is also needed for each primary care visit that occurred during the measurement period.

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

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

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Clinician Office/Clinic
If other:

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

2. Validity – See attached Measure Testing Submission Form

NQF_COC_testing_attachment_resbumitted_12-14-16-636174074345973370.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.3 For maintenance of endorsement
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

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

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

What were the statistical results of the analyses used to select risk factors?

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

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

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number
Measure Title: Continuity of Primary Care for Children with Medical Complexity
Date of Submission: 12/7/2016
Type of Measure:

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

Instructions

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

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

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

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

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

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

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

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

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

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

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record

☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

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

We obtained Medicaid Analytic eXtract (MAX) data for 17 state Medicaid agencies for 2005-2008.

1.3. What are the dates of the data used in testing? 1/1/2005-12/31/2008

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

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)
We obtained data from 17 states: Arizona, California, Illinois*, Indiana, Kansas, Kentucky, Louisiana*, Minnesota, Nebraska, New Jersey, New Mexico, New York, North Carolina*, Tennessee, Texas, Virginia, and Wisconsin. These 17 states were selected based on the quality of the Medicaid Analytic eXtract (MAX) data for research purposes.(Byrd & Dodd 2013)

*Only fee-for-service data were available for these states; all other states had both fee-for-service and managed care data.

Byrd & Dodd. 2013. Assessing the usability of MAX 2008 encounter data for comprehensive managed care. Medicare & Medicaid Research Review. 2008 Medicaid Managed Care Enrollment Report, Centers for Medicare and Medicaid Services, U.S. Department of Health and Human Services as of 06/30/2016; available at http://kff.org/medicaid/state-indicator/total-medicaid-mc-enrollment

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

A sample of individuals from the 17 states described in 1.5 were used in testing and analysis.

A total of 10,860,713 individuals in the 17 states used in our analyses were enrolled in Medicaid for the entire measurement year (2008) with no more than a 30-day gap. Of these, 1,095,068 were identified as having complex chronic conditions using the Pediatric Medical Complexity Algorithm, Version 2 (PMCA-V2). Of those, 11,438 met the age criterion of being >=1 year old and <=17 years old as of Dec. 31, 2008 and had at least 4 primary care visits in 2008. We note that 715,609 (6.6% of the source population) individuals had inadequate claims on which to run PMCA and so were not further evaluated for eligibility and were not included in any of our analyses (see Section 2b7). In sum, the 11,438 individuals that were identified as eligible for the measure were the individuals used in all of our analyses. We provide descriptive statistics for these individuals in Table 1.6.a below.

Table 1.6.a Demographics for individuals used in all analyses

	N	%
Total	11438	100
Child age	M=2.50, SD=3.30	
0-1*	7200	63
2-5	3164	28
6-12	587	5
13-17	487	4
Child gender		
Female	5352	47
Male	6086	53
Child race/ethnicity		
Asian/Pacific Islander	424	4
African American	1563	14
Hispanic	5435	48
White	3167	28
Other/multiracial	849	7
Lives in metropolitan statistical area (MSA)		
No	1731	15
Yes	9707	85
Temporary Assistance for Needy Families (TANF)/Low Income		
No	2193	19
Yes	9245	81
State		
AZ	478	4
CA	699	6
IL	1459	13
IN	858	8
KS	154	1
KY	242	2
LA	84	1
MN	64	1
NC	450	4
NE	84	1
NJ	171	2
NM	246	2
NY	1950	17

TN	469	4
TX	3730	33
VA	183	2
WI	117	1

*168 (1.5%) of the sample were 11 months old on December 31, 2008 and had no gaps in enrollment, so were retained in the analysis.

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

The same data were used for all aspects of testing.

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

Patient-level SDS variables that were available and analyzed in the data were child race/ethnicity, age, and gender; child eligibility for Temporary Assistance for Needy Families (TANF); and whether the child lives in a metropolitan statistical area (MSA).

2a2. RELIABILITY TESTING

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

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

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

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

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

To assess measure reliability, we calculated inter-unit (i.e., health plan-level) reliability, which refers to the degree to which measure scores are able to precisely distinguish between the performance levels of health plans, in this case state Medicaid agencies. We examined inter-unit reliability for the measure using intra-class correlations (ICCs) computed from individual scores. Specifically, we used a mixed effects logistic regression model with the individual measure score as the outcome (0/1) and random effects for health plan and obtained the ICC calculated from this model; Stata was used for this analysis. Secondly, we used this ICC to calculate the reliability of the measure based on the median number of eligible individuals per health plan using the Spearman-Brown formula (Allen and Yen, 1979). Specifically, this formula states that the expected reliability with N individuals per health plan is [N*ICC]/[1+(N-1)*ICC] where ICC is the value obtained from the mixed effects logistic regression described earlier. In addition, to look at this from a slightly different perspective, we also used the ICC and this formula to calculate the number of individuals needed per health plan (for future implementation) to ensure a reliability of 0.70,0.8 or 0.9. Re-arranging the Spearman-Brown formula shows that the sample size per health plan needed to ensure a reliability of e.g. 0.70 can be calculated as [0.70*(1-ICC)]/[ICC*(1-0.7)]. When entities such as health plans are being compared, measure inter-unit reliability of at least 0.70 is commonly considered adequate and reliability in the 0.80-0.90 range are considered excellent (Hargraves, Hays & Cleary, 2003; Nunnally 1994).

Allen M, Yen W. Introduction to Measurement Theory. Monterey, CA: Brooks/Cole;1979.
Nunnally JC, Bernstein IH. Psychometric Theory. New York: McGraw Hill; 1994.
Hargraves, L.J., Hays, R. D., & Cleary, P. D. (2003). Psychometric properties of the consumer assessment of health plans study (CAHPS®) 2.0 adult core survey. *Health services research*, 38(6p1), 1509-1528.

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

The intra-class correlation for health plans (in this case, state Medicaid agencies) was 0.28 (95% CI 0.16-0.44). The median number of eligible individuals per health plan in our test data was 246 (minimum 64, maximum 3,730) and thus the reliability of the measure is 0.98 indicating that there is a high degree of variability between health plans, compared to within states in terms of performance on this measure. For any health plan where this measure is used, the number of patients per health plan needed for reliability of 0.7, 0.8, and 0.9 are shown in Table 2a2.3.a below.

Table 2a2.3.a Reliability results

	Expected Reliability[1]		
	0.7	0.8	0.9
Number of Eligible Individuals per Health Plan Needed to Achieve Expected Reliability	6	11	24

[1] using Spearman-Brown prediction formula

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

These results demonstrate that with a reliability of 0.98, this measure is highly reliable and can be used to distinguish between health plans in terms of performance. In addition, only 6 eligible individuals per health plan are necessary to result in adequate reliability (0.7) of the measure. High reliability of 0.9 can be achieved with 24 eligible individuals per health plan. This should be feasible, at least at the state Medicaid agency level, since we observed a median of 246 eligible individuals per state Medicaid agency in our analysis (range of 64-3,730). In summary, these results demonstrate that we expect this measure to be able to reliably distinguish between health plans in terms of performance.

2b2. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

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

PERFORMANCE MEASURE: EMPIRICAL VALIDITY TESTING

At the individual level, we examined the relationship between the COC quality measure score and the occurrence of at least one ED visit in the measurement year. We used mixed-effects logistic regressions that predicted the occurrence of at least one ED visit from the measure score, with random intercepts for states. We examined models both unadjusted

and adjusted for characteristics associated with both the COC quality measure and the occurrence of at least one ED visit: child age, residence in a metropolitan statistical area (MSA), and eligibility for Temporary Assistance for Needy Families (TANF). This analysis is designed to look for evidence of convergent validity – that is, that higher COC measure scores are related to lower odds of an ED visit.

PERFORMANCE MEASURE: SYSTEMATIC ASSESSMENT OF FACE VALIDITY—DELPHI PANEL

The Delphi Panel process began with the nomination of 20 individuals by 10 stakeholder organizations including the American Academy of Pediatrics, the Academic Pediatric Association, the Society for Hospital Medicine, the Children's Hospital Association, the Medicaid Medical Directors Learning Network, Family Voices, the American Academy of Child and Adolescent Psychiatry, the Society for Adolescent Medicine, the National Association of Pediatric Nurse Practitioners, and the Society for Developmental and Behavioral Pediatrics. Nine of the nominees agreed to be members of our multi-stakeholder Delphi panel. All panelists were people deemed by the nominating organizations to have substantial expertise and/or experience related to care coordination for CMC (see Ad.1 for a list of panel members). The panel read the literature reviews written by project staff and reviewed and scored each proposed quality measure on validity. This method is a well-established, structured approach to measure evaluation that involves two rounds of independent panel member scoring, with group discussion in between. (Brook 1994) After reviewing literature reviews and draft quality measures, panel members were asked to rate each measure's validity on a scale from 1 (low) to 9 (high). Validity was assessed by considering whether there was adequate scientific evidence or expert consensus to support its link to better outcomes; whether there would be health benefits associated with receiving measure-specified care; whether they would consider providers who adhere more consistently to the quality measure to be providing higher quality care; and whether adherence to the measure is under the control of health care providers and/or systems. For a quality measure to move to the next stage of measure development, it had to have a median validity score > 7 (1-9 scale) and be scored without disagreement based on the mean absolute deviation from the median after the second round of scoring. This process ensures that only measures widely judged to be valid moved forward into measure specification.

Brook RH. The RAND/UCLA appropriateness method. In: McCormick KA, Moore SR, Siegel RA, eds. Clinical practice guidelines development: methodology perspectives. Rockville, MD: Agency for Health Care Policy and Research; 1994.

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

PERFORMANCE MEASURE: EMPIRICAL VALIDITY TESTING

The COC quality measure was significantly associated with decreased odds of having one or more ED visits ($p<0.05$). Table 2b2.3.a below shows the odds ratio (OR) and 95% confidence limits of an ED visit for children with a Bice-Boxerman COC index of 0.5 or greater (COC quality measure pass) compared to children with a Bice-Boxerman COC index below 0.5 (COC quality measure fail). Results are presented for models both unadjusted and adjusted for characteristics associated with both the COC quality measure and the occurrence of at least one ED visit: child age, residence in a metropolitan statistical area (MSA), and eligibility for Temporary Assistance for Needy Families (TANF). The unadjusted odds ratio was 0.90 ($p=0.0110$) and the adjusted odds ratio was 0.88 ($p=0.0021$).

We do not recommend risk adjustment of the COC quality measure for SDS when it is used to compare health plan performance. As described in the testing attachment section 2b2.2, we fit validation models both unadjusted and adjusted for SDS variables that were associated with both the COC quality measure and the validation metric (occurrence of at least one ED visit) in our test dataset. We fit the adjusted validation model to remove potential

confounders of the relationship between the COC quality measure and the validation metric in our particular test dataset. Results from regressions unadjusted and adjusted for SDS are presented in Table 2b2.3.a.

Table 2b2.3.a Empirical validity results

	Odds ratio (OR)	Lower Limit of 95% confidence interval (LCL)	Upper Limit of 95% confidence interval (UCL)	p-value
Unadjusted				
Bice-Boxerman COC index of 0.5 or greater (COC quality measure pass)	0.90	0.83	0.98	0.0110
Adjusted for SDS				
Bice-Boxerman COC index of 0.5 or greater (COC quality measure pass)	0.88	0.81	0.95	0.0021

PERFORMANCE MEASURE: SYSTEMATIC ASSESSMENT OF FACE VALIDITY—DELPHI PANEL

The scores for this measure from the 9 members of the panel after round 2 of Delphi scoring (scoring done after discussions at the in-person meeting) are presented in Table 2.b2.3.b below.

Table 2b2.3.b Delphi Panel results

	Median score (Scale 1-9)	Mean absolute deviation from median	Agreement status*
Face Validity	7.0	0.7	Indeterminate
Feasibility	7.0	0.6	Agreement

*This is a statistical assessment of whether panelists agreed, disagreed, or if status was indeterminate.

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

PERFORMANCE MEASURE: EMPIRICAL VALIDITY TESTING

These results indicate that the individuals with a Bice-Boxerman COC index >= 0.5 (i.e., pass the COC quality measure) had a significantly lower odds of having an ED visit in the measurement year. These results demonstrate the convergent validity of the COC quality measure with this validation metric.

PERFORMANCE MEASURE: SYSTEMATIC ASSESSMENT OF FACE VALIDITY—DELPHI PANEL

The results from the Delphi panel show reasonable levels of face validity. We note that the median score for validity was 7.0 on the scale of 1-9 where 7-9 is considered “valid”. In addition, the measure was scored without disagreement according to the mean absolute deviation from the median which also supports its face validity (Brook, 1994).

Brook RH. The RAND/UCLA appropriateness method. In: McCormick KA, Moore SR, Siegel RA, eds. Clinical practice guidelines development: methodology perspectives. Rockville, MD: Agency for Health Care Policy and Research; 1994.

_____

2b3. EXCLUSIONS ANALYSIS

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

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

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

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

_____

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

N/A. This is a structure quality measure.

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

☐ **No risk adjustment or stratification**

☐ **Statistical risk model with** 3 **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

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

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

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

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

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

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

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

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

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

2b4.9. Results of Risk Stratification Analysis:

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

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

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

We tested for differences in performance across the 17 state Medicaid agencies using a chi-square omnibus test. We then used logistic regressions to perform individual comparisons between each state's performance and the performance of the group as a whole.

Performance variation across sociodemographic characteristics was assessed using bivariate logistic regression (each characteristic was modeled without adjusting for other covariates).

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

Table 2b5.2.a Differences in COC Quality Measure Scores by State Medicaid Agencies

	Denominator	Numerator	COC Quality Measure Pass Rate (%)	P-value for omnibus test	Difference from overall mean of others	P-value for difference from overall mean of others

17 States pooled	11438	7583	66.30	<.0001	--	--
AZ	478	333	69.67	--	3.52	0.1118
CA	699	513	73.39	--	7.55	<.0001
IL	1459	1113	76.29	--	11.44	<.0001
IN	858	785	91.49	--	27.24	<.0001
KS	154	89	57.79	--	-8.62	0.0254
KY	242	77	31.82	--	-35.22	<.0001
LA	84	65	77.38	--	11.15	0.0334
MN	64	51	79.69	--	13.43	0.0263
NC	450	432	96.00	--	30.92	<.0001
NE	84	80	95.24	--	29.15	<.0001
NJ	171	135	78.95	--	12.84	0.0005
NM	246	90	36.59	--	-30.36	<.0001
NY	1950	1272	65.23	--	-1.29	0.2715
TN	469	256	54.58	--	-12.21	<.0001
TX	3730	2136	57.27	--	-13.40	<.0001
VA	183	129	70.49	--	4.26	0.2269
WI	117	27	23.08	--	-43.65	<.0001

Table 2b5.2.b Differences in COC Quality Measure Scores by Sociodemographic Characteristics

	COC Quality Measure Pass Rate (%)	p-value
Child age		
0-1 (reference group)	67.40	
2-5	65.17	0.0264
6-12	62.86	0.0246
13-17	61.40	0.0065
Child gender		
Female (reference group)	66.48	
Male	66.14	0.6974
Child race/ethnicity		
Asian/Pacific Islander	69.81	0.2578
African American	64.81	0.1227
Hispanic	66.20	0.4117
White (reference group)	67.07	
Other/multiracial	65.02	0.2609
Lives in MSA		
No (reference group)	69.27	
Yes	65.77	0.0046
TANF/Low Income		
No (reference group)	68.72	
Yes	65.72	0.0076

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

For the test assessing existing variation in the COC quality measure across more than one state Medicaid agency, we found that we were able to detect statistically and clinically meaningful differences in state Medicaid agency performance. We were also able to detect statistically and clinically meaningful differences in COC quality measure scores across demographic subgroups for the following characteristics: child age, residence in an MSA, and receipt of TANF.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

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

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

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

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

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

Individuals with inadequate claims on which to run PMCA were not able to be assessed for eligibility and thus were not included in our analyses. We examined the proportion of individuals this missing data represented and considered whether these missing data would be expected to bias our reliability and validity results.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing*

data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each)

A total of 715,609 (6.6% of the source population) individuals had inadequate claims on which to run PMCA and so were not further evaluated for eligibility and were not included in any of our analyses.

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

Given that this missing subset was a small proportion of our source population (6.6%), we expect that our results regarding reliability and validity would be robust to this missingness. Of course, there is the possibility that some bias may be present; however, this is a limitation of most analyses that use administrative data for measurement and testing.

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

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

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

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

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

3c. Data Collection Strategy

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

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Quality Improvement (external benchmarking to organizations) Quality Improvement (Internal to the specific organization) Not in use	

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

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

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

This measure is part of a set of care coordination quality measures the Center of Excellence on Quality of Care Measures for Children with Complex Needs (COE4CCN) developed as part of the Pediatric Quality Measurement Program, funded by AHRQ and CMS, using CHIPRA monies. It has not yet been implemented as the development and validation were just recently completed. The measure specifications, available online at the website http://www.seattlechildrens.org/research/child-health-behavior-and-development/mangione-smith-lab/measurement-tools/, are publicly available and non-proprietary, so interested parties can implement them at any time.

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

The Minnesota and Washington State Medicaid agencies have had representation on the National Advisory Board of our center (COE4CCN) since its inception. We will work with the National Academy for State Health Policy and the Medicaid Medical Directors Learning Network (MMDLN), which have both shown interest in enhancing the patient centered medical home in pediatrics. MMDLN has conducted quality improvement collaboratives focused on the patient-centered medical home. Given the high feasibility of implementing the proposed measure, it is likely that these organizations will encourage state Medicaid agencies to implement it.

Improvement

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

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

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

Improving COC for CMC would contribute to lower ED utilization and ambulatory care-sensitive hospitalizations by ensuring timely delivery of care to prevent health crises. There are opportunities for improvement with respect to both of these outcomes. In 2014, a 12-state study found that 32% of CMC had >1 ED visit that did not result in admission, which suggests that some of these ED visits could have been avoided.(1) In the same population, 12% of hospitalizations of CMC were for ambulatory-care sensitive conditions. The proposed quality measure on Continuity of Primary Care for Children with Medical Complexity would enable state Medicaid agencies to assess COC and ensure that COC is improved for this vulnerable population.

1. Berry JG, Hall M, Neff J, et al. Children with medical complexity and Medicaid: spending and cost savings. Health affairs (Project Hope). 2014;33(12):2199-2206.

4c. Unintended Consequences

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

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

None observed

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

None observed

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**

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

N/A; retrospective Medicaid Analytic eXtract (MAX) data from 2008 were used for field testing of the measure.

We selected states to include in this analysis based on the quality of the available data.(1)

References

1. Byrd & Dodd. 2013. Assessing the usability of MAX 2008 encounter data for comprehensive managed care. Medicare & Medicaid Research Review.

2008 Medicaid Managed Care Enrollment Report, Centers for Medicare and Medicaid Services, U.S. Department of Health and Human Services as of 06/30/2016; available at http://kff.org/medicaid/state-indicator/total-medicaid-mc-enrollment

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

N/A; retrospective Medicaid Analytic eXtract (MAX) data from 2008 were used for field testing of the measure.

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**

Describe how feedback was obtained.

N/A; Feedback was not obtained due to the retrospective nature of the field test.

4d2.2. **Summarize the feedback obtained from those being measured.**

N/A; Feedback was not obtained.

4d2.3. **Summarize the feedback obtained from other users**

N/A; not currently in use by others.

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
N/A

5. Comparison to Related or Competing Measures

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

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

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

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

5a. Harmonization of Related Measures
The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

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

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** PMCA_ICD9_DX_to_ICD10_Mapping_DRAFT_December_2016.xlsx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Seattle Children´s Research Institute
Co.2 **Point of Contact:** Rita, Mangione-Smith, rita.mangione-smith@seattlechildrens.org, 206-884-8242-
Co.3 **Measure Developer if different from Measure Steward:** Seattle Children´s Research Institute
Co.4 **Point of Contact:** Rita, Mangione-Smith, rita.mangione-smith@seattlechildrens.org, 206-884-8242-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
WORKGROUP MEMBERS:

1. Rita Mangione-Smith, MD, MPH; Seattle Children's Research Institute/ University of Washington, Seattle, WA; Oversaw entire project (study PI), including literature reviews, measure development, Delphi panel, measure specification, field testing, and analysis.
2. Kimberly Arthur, MPH; Seattle Children's Research Institute, Seattle, WA; Literature review and measure development
3. Courtney Gidengil, MD, MPH; Boston Children's Hospital/ Harvard Medical School/ RAND Corporation, Boston, MA; Literature review, measure development, analytic team
4. Eric Schneider, MD, MSc; RAND Corporation, Boston, MA (now Commonwealth Fund); Provided oversight and participated in all aspects of measure development and testing
5. Elizabeth McGlynn, PhD; Center for Effectiveness and Safety Research, Kaiser Permanente, Pasadena, CA; Provided oversight and participated in all aspects of measure development and testing
6. Layla Parast, PhD; RAND Corporation, Santa Monica, CA; Biostatistician and analytic team lead
7. Q Burkhart, MS; RAND Corporation, Santa Monica, CA; Data analyst and analytic team
8. Marc Elliott, PhD; RAND Corporation, Santa Monica, CA; Biostatistician and analytic team
9. Hangsheng Liu, PhD; RAND Corporation, Boston, MA; Biostatistician and analytic team
10. Scott Ashwood, PhD; RAND Corporation, Santa Monica, CA; Data analyst and analytic team
11. Julie A. Brown; RAND Corporation, Santa Monica, CA; Survey design and data collection
12. Laurie Cawthon, MD, MPH; WA State Department of Social and Health Services, Olympia, WA; Field testing, data acquisition and analysis

DELPHI PANEL MEMBERS:

1. Richard Antonelli, MD, MS
Medical Director of Integrated Care and Strategic Partnerships
Medical Director Physician Relations and Outreach
Boston Children's Hospital
Assistant Professor of Pediatrics
Harvard Medical School
Nominated by American Academy of Pediatrics (AAP)

2. Allison Ballantine, MD, MEd
Assistant Professor of Pediatrics
University of Pennsylvania School of Medicine
Section Chief of Education
Medical Director, Integrated Care Services
Division of General Pediatrics
Attending Physician Palliative Care Team
Attending Physician Inpatient General Pediatrics
The Children's Hospital of Philadelphia
Nominated by Society of Hospital Medicine (SHM)

3. Jennifer Bolden-Pitre, MA, JD
Director of Integrated Systems,

Statewide Parent Advocacy Network
Family Fellow, Leadership Education in Neurodevelopmental Disabilities
Children´s Hospital of Philadelphia
Nominated by Family Voices

4. Carol A. Ford, MD
Professor of Pediatrics
Orton Jackson Endowed Chair in Adolescent Medicine
University of Pennsylvania
Chief, Craig Dalsimer Division of Adolescent Medicine
The Children´s Hospital of Philadelphia
Nominated by Society for Adolescent Health & Medicine (SAHM)

5. Jason Kessler, MD, FAAP, CHBE
Medical Director
Iowa Medicaid Enterprise
Nominated by Medicaid Medical Directors Learning Network (MMDLN)

6. Karen Kuhlthau, PhD
Associate Professor, Pediatrics
Harvard Medical School
Associate Sociologist, Pediatrics
Center for Child and Adolescent Health Policy
Massachusetts General Hospital for Children
Nominated by Academic Pediatric Association (APA)

7. Dennis Kuo, MD, MHS
Assistant Professor of Health Policy and Management
Fay W. Boozman College of Public Health,
University of Arkansas for Medical Sciences
Assistant Professor of Pediatrics
Section on General Pediatrics
Center for Applied Research and Evaluation,
University of Arkansas for Medical Sciences
Pediatrician
Medical Home Program for Children with Special Needs,
Arkansas Children's Hospital
Nominated by Children's Hospital Association (CHA)

8. Wendy Sue Looman, PhD, RN, CNP
Pediatric Nurse Practitioner
Cleft Palate and Craniofacial Clinic
School of Dentistry, University of Minnesota
Associate Professor
School of Nursing, University of Minnesota
Nominated by National Association of Pediatric Nurse Practitioners (NAPNAP)

9. Karen Pierce, MD, FAPA, FAACAP
Attending Physician
Department of Child and Adolescent Psychiatry
Children's Memorial Hospital, Chicago, Illinois
Clinical Associate Professor
Feinberg School of Medicine, Northwestern University Medical School
Department of Psychiatry and Behavioral Sciences
Nominated by American Academy of Child & Adolescent Psychiatry (AACAP)

Measure Developer/Steward Updates and Ongoing Maintenance Ad.2 **Year the measure was first released:** Ad.3 **Month and Year of most recent revision:** Ad.4 **What is your frequency for review/update of this measure?** Ad.5 **When is the next scheduled review/update for this measure?**
Ad.6 **Copyright statement:** Ad.7 **Disclaimers:**
Ad.8 **Additional Information/Comments:**

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3154
Measure Title: Informed Coverage
Measure Steward: The Children´s Hospital of Philadelphia
Brief Description of Measure: Improved measurement of the continuity of insurance coverage in the Medicaid and CHIP population is needed to help maximize insurance continuity and coverage for vulnerable children. To further this goal, the AHRQ-CMS CHIPRA PQMP Center of Excellence at the Children's Hospital of Philadelphia developed the metric Informed Coverage. The metric is designed to more accurately measure coverage among children enrolled in Medicaid or CHIP at the state level and overcome the current inability in the Medicaid Analytic eXtract (MAX) dataset to determine whether a child disenrolled from Medicaid and CHIP due to loss of eligibility (such as due to parental income increase or the acquisition of employer-sponsored insurance, a "good" reason) or failure to appropriately re-enroll (a "bad" reason). This measure can help federal and state programs develop strategies to retain children eligible for coverage and minimize gaps that can occur during the renewal process. Informed Coverage assesses the continuity of enrollment of children in publicly financed insurance programs (Medicaid and CHIP), as defined by the ratio of enrolled month to eligible months over an 18 month observation window. Informed Coverage uses a natural experiment based on the random event of appendicitis to "inform" the estimate of coverage in a given state, bounded by two extreme assumptions regarding unknown eligibility information: Coverage Presumed Eligible (PE) and Coverage Presumed Ineligible (PI).
Developer Rationale: States are frequently asked to determine public insurance participation rates or measure continuity of enrollment among vulnerable children, both for federal compliance audits and performance-based incentives, and for internal studies concerning vulnerable populations (Patrick et al., 2012; Daly 2003; National Conference of State Legislatures Health Policy Tracking Service, 2003). Participation rates are defined as the fraction of eligible children who are enrolled (Kenney et al., 2009). We developed and validated this administrative claims–based participation metric, "Informed Coverage," using a naturally occurring randomization observed inside each state that dynamically informs assumptions about patterns of eligibility and allows statewide estimates of participation rates using only administrative claims data. This standardized measure can be used by states as a potential indicator of quality and access. The issue of enrollment and retention is a long-standing concern for publically financed insurance programs, and one that states have likely examined using less formal means. Because Medicaid/CHIP enrollees are from low-income families, this measure will benefit vulnerable children it will hold states accountable for retaining children eligible for public coverage. Where data capacity permits, this measure also takes into account children switching from Medicaid and CHIP and vice-versa instead of treating children as disenrolled from public insurance.

Numerator Statement: The numerator for Informed Coverage represents the sum (within a state) of months enrolled in Medicaid/CHIP for all children over an 18-month window.
Denominator Statement: The sum (within a state) of months eligible for Medicaid/CHIP for all children (0-18 years) over an 18-month window. In addition, months that could be defined as "eligible" are based on known events recorded in the MAX data that would affect eligibility (birth or ageing out).

S.8. **Denominator Exclusions:** For the appendicitis calculation, the population is limited to children between the ages of 2 to 16 years old. To determine what is the best assumption to use (either the Appendectomy Coverage Rate (or ACR), PI, or PE) inside each state, we compare the observed appendectomy coverage rate in a state, to the estimated coverage rate that would be calculated in that state with either PI, or PE assumptions.

Measure Type: Outcome
Data Source: Claims (Only)
Level of Analysis: Population : Regional and State

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary

This is a state/regional-level outcome measure, for which the measure developer conducted a literature review related to effects of uninsurance and gaps in coverage in the United States. The developer provides the following rationales:

- *Potential for Quality Improvement*: State programs have an interest in retaining eligible children and preventing inappropriate breaks in coverage, many of which occur during the renewal process.
- *Prevalence of Issue*: Studies have used the Medical Expenditure Panel Survey, the National Health Interview Survey, and the National Survey of Children's Health have found that prevalence of uninsurance and unstable coverage among children in the United State is between 9 to 11.1 million for children with gaps in coverage and 5 to 6 million for those with no insurance in a given survey year (Bethell, 2011; Satchell, 2005).
- *Fiscal Burden of Lack of Insurance and Gaps in Coverage*: Disenrollment and reenrollment establishes additional administrative costs. In addition, disenrollment would increase the costs of health care due to increased ED visits and hospital stays.
- *Child Health Outcomes*: A regular source of care allows for treatment of chronic health conditions, provides routine preventive care, and management of acute and urgent problems (Olson, 2005). Continuous coverage ensures that children and adolescents can receive continuity of care without gaps and permits children's health conditions to be monitored regularly and treatments adjusted to maximize health and prevent exacerbations or worsening of conditions that might lead to hospitalization (Fairbrother, 2004; Weissman, 1992). Finally, continuity of coverage may allow time for greater engagement with clinicians in treatment decisions that lead to greater satisfaction with services and better health status (Holl, 2000; Kenney, 2007; Shone, 2005).
- *Policy Factors*: Many states have implemented various strategies for streamlining and simplifying the enrollment and renewal processes for Medicaid and CHIP. Generating additional evidence that elucidates the pathway that encompasses policy context, insurance coverage, service delivery, and outcome is important for policy decisions.

Question for the Committee:

- *Is there at least one thing that measured entities can do to achieve a change in the measure results?*

Guidance from the Evidence Algorithm: Health outcome (Box 1)→ Relationship between outcome and provider action (Box 2)→ Pass

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

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

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

The developer reports the following gap information:

- The Informed Coverage fraction distributions across states for the time period January 1, 2008 to June 30, 2009 are presented in Table 1 of the Appendix, as are the distributions for the intermediate calculations.
    - The mean state Informed Coverage value was 0.7949 (SD 0.1035).
    - The minimum state Informed Coverage Value was 0.3814, the maximum was 0.9350, and the IQR was 0.7474, 0.8580.
    - The deciles of the state Informed Coverage values from 10% through 90% were as follows: 0.6623, 0.7326, 0.7576, 0.7778, 0.8261, 0.8469, 0.8571, 0.8618, 0.8840.
- Data quality limited analyses to 43 states. Six states failed to have sufficient reporting of managed care claims for the data utilized in the development of this measure and thus were eliminated from analyses for this time period and dataset: Kentucky, Massachusetts, Mississippi, Ohio, Pennsylvania, and West Virginia. Additionally, Maine and the District of Columbia were found to have excessive quality issues in their inpatient records and were eliminated.

Disparities:
The developer states that "Disparities in continuity of coverage according to ethnicity, geography, insurance type, and special health care need have been observed throughout the literature," and notes the following:

- Race/ethnicity (Appendix Table 4): The developer stratified the 18-month informed coverage fraction by enrollee race/ethnicity. Coverage fractions varied by race within and across states with variations in the race/ethnicity groups with the highest and lowest Informed Coverage values within the state.
- Special Health Care Needs (Appendix Table 5): The developer stratified the 18-month informed coverage fraction by enrollee chronic condition status. Informed Coverage value were generally higher for children with chronic care needs.
- Socioeconomic Status (Appendix Table 6-8): The coverage fraction across the poverty quartiles lacked a coherent pattern, although the extremes show that ZIP codes with a lower percentage of enrollees below the FPL had better coverage than ZIP codes with a higher percentage above the FPL. Differences across income quartiles were also small and lacked a coherent pattern across states. When looking at just the extremes the lowest income quartile always had a better coverage than those in the highest quartile. Trends for education were clear: while coverage fractions were generally homogenous and never differed by more than ten percentage points between the most- and least-education quartiles, coverage fractions in every state improved as the high school graduation rate of enrollee ZIP codes fell.
- Rurality/Urbanity (Appendix Table 9): The Informed Coverage values were drastically lower for enrollees who lacked a geographic status (zip code information missing), compared to any other category. Informed Coverage values were generally similar between urban, rural, and urban cluster areas in each state.

Question for the Committee:
- *Is there a gap in care that warrants a national performance measure?*

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

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

- Evidence provided suggests that a large number of children experience unstable insurance coverage in any given year. Disenrollment and reenrollment is costly to states administratively, and results in increased hospital use and ED visits for these children due to a lack of continuity in care. Regular continuous care allows for children's' conditions to be monitored and treated preventing the exacerbation of conditions and costly hospitalizations.
- This is a State-level outcome measure with strong evidence linking this outcome to improved health. States can take actions to improve continuous enrollment, including presumptive eligibility, changing the time intervals between recertification and improving outreach efforts.

- The performance data provided showed that the Informed Coverage value was higher for children with chronic conditions. Persons within the lowest income bracket had higher coverage than those with in the highest income bracket. Those without missing geographic information had lower coverage than those identifying a geographic location. While there appears to be some variation within the measure that could point to policy interventions to improve coverage, the gaps uncovered were not surprising or very drastic.
- Data was presented to support significant variation in continuous enrollment across the country, as well as between groups of different race and ethnicity.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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

Data source(s): Administrative claims
Specifications:

- Level of analysis: Population: Regional and State
- Interpretation of score: high score=better quality
- Numerator: *The numerator for Informed Coverage represents the sum (within a state) of months enrolled in Medicaid/CHIP for all children over an 18-month window.*
    - In addition, months that could be defined as "eligible" are based on known events recorded in the MAX data that would affect eligibility (birth or ageing out). A month is considered "covered" if a child has greater than 14 enrolled days in that month or if there is an indicator for S-CHIP coverage for that month.
    - This sum is derived using one of three methods based on the state. Children are either presumed eligible (PE), presumed ineligible (PI), or a state's Appendectomy Coverage Ratio (ACR) is used. To determine what is the best assumption to use (either the Appendectomy Coverage Rate (or ACR), PI, or PE) inside each state, compare the observed appendectomy coverage rate in a state, to the estimated coverage rate with either PI, or PE assumptions. If PE < ACR < PI, utilize ACR. If ACR > PI, use PI, and if ACR < PE use PE. Appendicitis is defined using principal diagnosis or procedure using ICD-9 and -10 codes.
- Denominator: *The sum (within a state) of months eligible for Medicaid/CHIP for all children (0-18 years) over an 18-month window. In addition, months that could be defined as "eligible" are based on known events recorded in the MAX data that would affect eligibility (birth or ageing out).*
- Exclusions: For the appendicitis calculation, the population is limited: children < 2 years and >16 years are excluded.
- A calculation algorithm is provided.
- The developer states that stratification can be performed for Informed Coverage using any desired strata that policymakers choose to study—e.g., stratification can be performed within states based on the type of Medicaid and CHIP programs, or by race. The measure does not include required risk adjustment or risk stratification. The developer states that risk adjustment is not appropriate, because "To perform any risk adjustment would evaluate states using different standards, granting leeway for some states. As all states should strive to improve their process to ensure enrollment of eligible children, risk adjusting would not provide an accurate picture of the state's need to improve its coverage status."
- There is no sampling for this measure.

This measure uses the random event of appendicitis to inform the estimate of coverage in a given state. Appendicitis is utilized in calculating the measure as it:

- Has an acute onset (reflecting a discrete point in time);

- Has an incidence rate that is not influenced by prior care, insurance coverage, or by factors that may influence obtaining coverage, such as socioeconomic status; and,
- Requires hospitalization for all children regardless of insurance status.
    - If a child is hospitalized and generates a bill seen in the Medicaid claims, he/she must have been eligible for Medicaid. If a child was not enrolled at the time of developing appendicitis, but was eligible, the appendicitis should still be observed because Medicaid and most CHIP programs allow up to 3 months of retroactive coverage and most states have policies of presumptive eligibility for their public insurance program. The numerator for the appendicitis calculation is the number of children with an appendicitis hospitalization during the same 18-month observation window used for the Coverage PE and Coverage PI intermediate calculations, who are enrolled in Medicaid/CHIP four months prior to their inpatient stay.

Questions for the Committee:

- *Is the natural experiment of appendectomy an appropriate surrogate for calculating informed coverage?*
- *Is the way PE, PI, and ACR defined appropriate? Is the way the calculation method is chosen for each state appropriate?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing, Testing attachment

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

SUMMARY OF TESTING

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

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

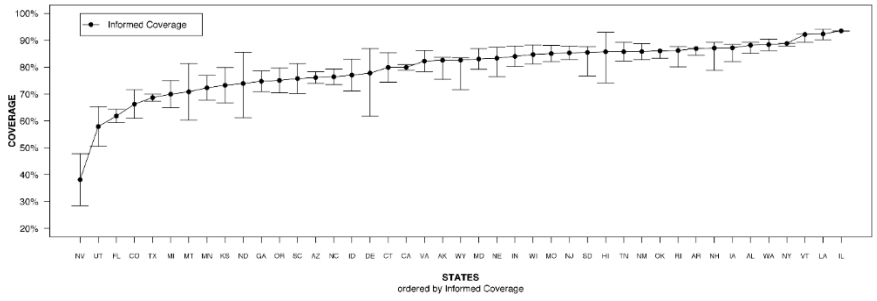
Method(s) of reliability testing:

- State Informed Coverage values for January 2008 – June 2009 were calculated using the weighted mean of the state's Coverage PE and Coverage PI with the weights based on the intermediate calculation's closeness to the state Appendectomy coverage rate. Reliability was assessed by determining if states could be identified with non-overlapping 95% CI.

Results of reliability testing:

- The developer concludes that states were identified with non-overlapping Informed Coverage values, indicating that measure values are variable across the entities measured.

Informed Coverage (95% CI) by State ordered by State Informed Coverage for Development (Testing) Time Period



Questions for the Committee:

- *Is the method described appropriate to assess reliability of the measure?*
- *Do the results demonstrate sufficient reliability so that differences in performance between and among the measured entities can be identified?*

Guidance from the Reliability Algorithm

Submitted specifications precise, unambiguous and complete (Box 2)→ Empirical reliability testing conducted (Box 4)→ Method described and appropriate for assessing proportion of variability due to real differences (Box 5)→ Low certainty that performance measure scores are reliable (Box 6)→ Low

The highest possible rating is HIGH.

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

Rationale: The testing data identify a significant number (and possibly majority) of overlapping confidence intervals around the mean for the 43 states. Given NQF's endorsement focus is on accountability purposes (public reporting and payment), the ability of a measure to reliably distinguish among the measured entities is a key attribute (in contrast to using the measure to guide quality improvement efforts).

2b. Validity

2b1. Validity: Specifications

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

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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

SUMMARY OF TESTING

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

Method of validity testing of the measure score:

☐ **Face validity only**

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

Validity testing method:

- To examine construct validity, the developer calculated Pearson correlations and median absolute errors between the external gold standard of the American Community Survey (ACS) metric using data from 2008-2010. Construct validity of Informed Coverage also was examined by comparison with two other measures:
  1) Duration of first observed enrollment, defined as defines a period of enrollment that begins after new entry to Medicaid/CHIP, birth, or an insurance gap of at least one month, and reports percentage still enrolled at 6, 12, or 18 months.
  2) Continuity Ratio, which calculates the average number of children enrolled per month divided by the number of children enrolled at any point in the year.
- Bland-Altman plots were examined for ACS versus Informed Coverage and ACS versus the Continuity Ratio. It was hypothesized that Informed Coverage would have a higher correlation with the ACS than the other two measures.

Validity testing results:

- Using the ACS survey as a gold standard, the developer states that participation rates using Informed Coverage were highly correlated with the survey and error rates were low. Informed Coverage was well correlated with the ACS metric across the 43 states (r = 0.86 (0.76, 0.92)), and showed similar correlation with ACS in the validation set (r = 0.86 (0.76, 0.92)). The Continuity Ratio and Duration displayed far lower correlations in the two time periods.

- **Table 1: Pearson correlations** *(with 95% CIs) between metrics across states (N=43) in two 18-month time periods, January 2008-June 2009, used with the 2008 ACS and a split sample of the 2009 ACS for metric development, and July 2009-December 2010, used with the remainder of the 2009 ACS split sample and the entire 2010 survey, for validation.*

Table 2. Pearson Correlation Results				
		Continuity Ratio	**Duration**	**ACS**
Informed Coverage	**Development**	0.77[d] (0.61,0.87)	0.52[c] (0.26,0.71)	0.86[d] (0.76,0.92)
	Validation	0.73[d](0.55,0.85)	0.49[c] (0.23,0.69)	0.86[d] (0.76,0.92)
Continuity Ratio	**Development**	1	0.83[d] (0.70,0.90)	0.75[d] (0.57,0.85)
	Validation		0.71[d] (0.53,0.83)	0.67[d] (0.46,0.80)
Duration	**Development**		1	0.46[b] (0.19,0.67)
	Validation			0.51[c] (0.24,0.70)
[a] p<0.05, [b] p<0.01, [c] p<0.001, [d] p<0.0001				

- The median absolute error (MAE) between ACS and Informed Coverage was calculated. In the development data, the Informed Coverage showed a low MAE of 3.29% with respect to the 2008/2009 ACS survey and was 3.67% in the validation study. The MAE between the ACS and the Continuity Ratio was more than twice as large than for Informed Coverage for both the development and validation data sets.

Questions for the Committee:

- *Were the correlation metrics and hypotheses used for assessing validity (ACS, Duration, and Continuity Ratio) appropriate to assess validity of this measure?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

- The measure encompasses 0 to 18 years, but only children between the ages of 2 and 16 years are included in the intermediate calculation of the Appendicitis Coverage rate.
  - In the justification for the exclusions for the ACR, the developer states that a diagnosis of appendicitis in children 2 years or less is "incredibly rare and unreliable."
  - The developer states children 16 years or older were excluded because of concern for censoring due to aging out during the 18-month observation window. The developer states that it "did not believe that including 16 to 18 year olds could possibly change the coverage rates to any appreciable amount."
- In its analysis of the exclusions for the ACR:
  - There were 32,653 inpatient claims for appendicitis or appendectomy in children 0-18 years old between January 1, 2008 and June 30, 2009, of these 5,079 (15.6%) were excluded from the Appendicitis Coverage rate calculation because the patient was either between 2 or less or 16 years or older.

- The developer reports a "minimal difference" (Median: 0.61%; Range: -2.09% to 2.55%) in the intermediate calculation of the Appendicitis Coverage rate when comparing the value calculated when using patients between the ages of 2-16 and patients between the ages of 0-18 years old.

Questions for the Committee:

- *Are the exclusions consistent with the evidence?*
- *Are any patients or patient groups inappropriately excluded from the measure?*
- *Does the Committee concur that the impact of excluding <2 and >16 years results in a "minimal" difference given the other empirical testing data?*

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

Rationale for no risk adjustment:

- The developer states that "States need to know whether their participation rates are above or below the participation rates of other states and act accordingly. To perform any risk adjustment would evaluate states using different standards, granting leeway for some states. As all states should strive to improve their process to ensure enrollment of eligible children, risk adjusting would not provide an accurate picture of the state's need to improve its coverage status. States need to adjust their approach to enrolling children based on unadjusted measurements of the problem. Therefore, it is not our intent to risk adjust these rates."

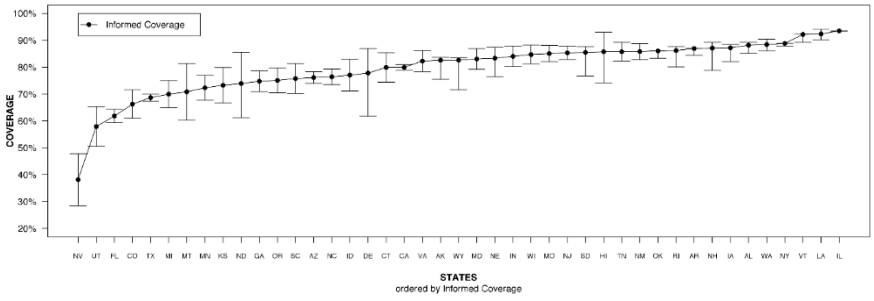
Questions for the Committee:

- *Is the rationale for not including a risk-adjustment strategy in the measure appropriate?*
- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*

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

- The developer determined that states could be identified by non-overlapping 95% confidence intervals (Table 1) and concluded there are practical, meaningful differences across measured entities.
- The developer also noted the rates for the Informed Coverage Value was 0.3814 to 0.9350, and the inter-quartile ranges was 0.7474, 0.8580.
- Figure 1 graphs performance by state.

Informed Coverage (95% CI) by State ordered by State Informed Coverage for Development (Testing) Time Period



Question for the Committee:

- *Does this measure identify meaningful differences in quality between and among measured entities?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

The developer provided the following information related to missing data and noted states are excluded, as necessary, based on defined parameters:

- Mathematica Policy Research reports for 2008, 2009, and 2010 MAX data report the following anomalies for data needed for the Informed Coverage measure (Overarching Note: DC, HI, MO, ND, PA, UT, and WI were not included in MAX 2008 because the corresponding MSIS files were unavailable or contained significant data problems.)
    - Values greater than 2.0 percent of records with no reported Medicaid enrollment (excluding S-CHIP only enrollees) are above the expected level and are considered anomalous. The following states were found to have data anomalies for 2008 enrollment data: Arizona (21.6%), California (2.8%), Florida (2.1%), Hawaii (3.4%), Michigan (3.7%), Oklahoma (3.6%), Utah (3.6%), and Washington (2.1%). The following states were found to have data anomalies for 2009 enrollment data: California (2.6%), Hawaii (3.5%), Utah (3.9%), and Washington (2.3%). The following states were found to have data anomalies for 2010 enrollment data: Arizona (21.6%), California (2.8%), Florida (2.1%), Hawaii (3.4%), Michigan (3.7%), Oklahoma (3.6%), Utah (3.6%), and Washington (2.1%).
    - In a given year, the percent of records with a primary diagnosis included in the inpatient claims file is evaluated. None of the 43 states used in the analysis were found to have less than 100.0% of annual claims including a primary diagnosis. The following state was found to have less than 100.0% of 2010 claims including a primary diagnosis: Kansas (0.0%).
    - In a given year, the percent of records with any procedure code included in the inpatient file is evaluated. Values less than 50.0 percent are below the expected level and are considered anomalous. The following states were found to have less than 50.0% of 2008 claims including a procedure code: Arkansas (48.2%), Georgia (45.1%), and Rhode Island (0.2%). The following states were found to have less than 50.0% of 2009 claims including a procedure code: Arkansas (47.3%), Minnesota (50.0%), Missouri (43.2%), and Rhode Island (0.2%). The following states were found to have less than 50.0% of 2010 claims including a procedure code: Arkansas (49.0%), Connecticut (47.4%), Idaho (45.7%), Kansas (0.0%), Maryland (46.7%), Minnesota (49.8%), Missouri (42.6%), and Rhode Island (0.2%).
- For persons to be evaluated in the Informed Coverage measure, there must be evidence of enrollment within the 18-month observation period or the 18-month lookback period when conducting the presumed eligible calculation. Therefore, because there is no evidence of enrollment, these individuals would not be included at any level of the Informed Coverage measure.

Guidance from the Validity Algorithm
Specifications consistent with evidence (Box 2.)→ Potential threats to validity empirically assessed (Box 3)→ Empirical validity testing conducted using measure as specified (Box 6)→ Appropriate method (Box 7)→ Moderate confidence (Box 8c) → Low

The highest possible rating is HIGH.

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

Rationale: The meaningful difference data identify a significant number (and possibly majority) of overlapping confidence intervals around the mean for the 43 states. Given NQF's endorsement focus is on accountability purposes (public reporting and payment), the ability of a measure to reliably distinguish among the measured entities is a key attribute (in contrast to using the measure to guide quality improvement efforts).

Committee pre-evaluation comments

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

- Data elements appear to be clearly defined. ICD-9 and ICD-10 codes for Appendicitis are provided. No risk adjustment is necessary for the measure and justification is provided. This measure may not be consistently implemented across states and locales because of the lack of quality data available, reducing the measures usefulness to make national comparisons.

- Clear.
- Reliability testing was done on all states with available data. Reliability was assessed by determining if states could be identified with non-overlapping Informed Coverage Values, which according to the developers indicated that the measure values were variable across states. Values appear to be reliable.
- Done to the extent possible.
- The ACS, duration, and continuity ratio appear appropriate to assess validity of the Informed coverage metric. The Informed Coverage metric was highly corrected with the ACS for the two-time period and evidenced low error rates. This measure could be used as an indicator for states and regions to determine the quality of their efforts to maintain children's insurance coverage. Also, the measure could be an indicator of children's access to appropriate and necessary care. The variability in the scores between the states appears tight. Not certain how states could compare themselves when they are 1, 2, or 3% different.
- The exclusion of the >26 year olds would appear to miss a significant number of children. The exclusion of the <2 year olds appears appropriate given the conditions rarity in this group.

Criterion 3. Feasibility

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

The developer reports the following:

- All data elements are in defined fields in a combination of electronic sources and coded by someone other than person obtaining original information.
- Informed Coverage is designed to be used with the Medicaid Analytic eXtract (MAX) or similar administrative datasets. However, states and programs do not have consistent reporting standards when contributing to MAX. Some states do not report enrollment data, and no states reports claims for its state-funded (S-CHIP) programs.
- Although Medicaid and Medicaid-Expansion (M-CHIP) data are generally complete, some states do not report enrollment data.
- For children enrolled in Medicaid or M-CHIP, enrollment is reported as days per month (>14 days of enrollment is considered covered for that month). For states reporting S-CHIP, enrollment status is recorded monthly, via an S-CHIP indicator in the MAX data. In states that do not report S-CHIP enrollment to MAX, we must assess only the Medicaid and M-CHIP children.
- Six states were excluded from the testing sample after failing the following criteria: "In order to utilize the Informed Coverage metric, we must trust that states provide bills for children in managed care who develop appendicitis, but managed care claims are sometimes absent from MAX (Byrd and Verdier, 2011; Levinson, 2009). In order to address this problem, we review the managed care data reporting relative to fee-for-service (FFS) or primary care case management (PCCM) systems for inpatient appendicitis claims by state. In each state, over a calendar year, the managed care enrollment rate among children who had appendicitis was compared to that of children without appendicitis to assess whether claims for the managed care children were adequately reported. Each child with appendicitis was matched to 10 children without appendicitis via Mahalanobis distance optimal matching (Rosenbaum, 2010) with a distance matrix that included age and exact matched on gender, the two most clinically relevant risk factors for appendicitis (Addiss, 1990). This generated a control pool of children that had the same gender and very similar, if not identical, age to their matched counterpart. For each child with appendicitis, to avoid bias of retroactive coverage, a point-in-time four months before the date of appendicitis admission was used to determine whether the child was covered via FFS/PCCM or managed care, and the same month was used for their non-appendicitis matched counterpart. In the context of non-inferiority testing (Wellek, 2010), a state was deemed to have insufficient managed care appendicitis claims if the 95% confidence interval for the managed care rate in the appendicitis children minus the managed care rate in the matched controls was completely below -2%. Six states failed to meet this criterion and thus were eliminated from analyses for this time period and dataset: Kentucky, Massachusetts, Mississippi, Ohio, Pennsylvania, and West Virginia. Additionally, Maine and the District of Columbia were found to have excessive quality issues in

their inpatient records and were likewise eliminated. The results of this validation process are included in Appendix Table 10."

- The developer recommends that "states with separate Medicaid and CHIP administrations should develop ways to routinely merge their data to enhance the feasibility of the measure and facilitate implementation. In addition, routine inclusion of several specific elements (e.g., reason for enrollment, reason for disenrollment, English proficiency, etc.) would provide useful information."

Questions for the Committee:

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

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

Committee pre-evaluation comments

Criteria 3: Feasibility

- Some states do not consistently collect enrollment data. Some states that do collect this information do so inconsistently. Depending on the condition, there may be insufficient claims to conduct the analysis. Gaps in data collection make it difficult to compare across states. This study eliminated 12% of the states for this analysis due to missing data.
- None.

Criterion 4: Usability and Use

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

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

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

OR

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

- The developer indicates that if the measure is endorsed, it intends to work with the National Association of Medicaid Directors to implement the measure. It is unclear whether this would be for public reporting/accountability or internal quality improvement only.

Accountability program details: N/A

Improvement results:

- N/A: The measure is new and has not had widespread or targeted distribution to audiences of interest yet.

Unexpected findings (positive or negative) during implementation:

- N/A: No unintended consequences have been identified.

Potential harms: N/A

Vetting of the measure:

- The developer involved Medicaid/CHIP Programs from some states in its development, but the results, data, and assistance have not been provided to those states that are or may be implementing it.

Feedback: N/A

***Questions for the Committee**:*

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

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

Committee pre-evaluation comments

Criteria 4: Usability and Use

- If implemented consistently, this measure could be used to track access to health care for children, which can then lead to improvements in access and ultimately health outcomes. This is a new measure and not currently in use for accountability.
- Yes.

Criterion 5: Related and Competing Measures

Related or competing measures: N/A

Harmonization: N/A

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

RATIONALE IF NOT ELIGIBLE: The measure is not eligible because it has not been vetted by those being measured or other users.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 3154
Measure Title: Informed Coverage
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: 12/7/2016

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

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

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

Notes

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

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome
☒ Health outcome: Proportion of months with public insurance coverage of months eligible
☐Patient-reported outcome (PRO): Click here to name the PRO

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

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

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

The purpose of this measure is to determine the best method to estimate statewide public insurance participation rates in children using an informed approach based on a natural experiment of appendicitis. By states having a better estimate of participation in their Medicaid and CHIP programs, they can work to improve processes of ensuring enrollment of children that are eligible and preventing unintended interruptions in coverage. By doing so, states will be working towards of improving access and quality of care for their children.

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

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

Potential for Quality Improvement

We believe there are important policy implications for the Continuity of Insurance- Informed Coverage metric. State programs have an interest in retaining eligible children and preventing inappropriate breaks in coverage, many of which occur during the renewal process (Southern Institute, 2009). Many states have been engaged in efforts that have been shown to maximize continuous enrollment. These include streamlining and simplifying the enrollment and renewal processes for Medicaid and CHIP (Kaiser, 2012; Pati, 2012; Ku, 2013). Simplified enrollment procedures include express lane eligibility, Social Security Administration (SSA) data match to verify citizenship, and electronic forms. Simplified renewal procedures include using pre-populated forms and 12-month continuous eligibility. States where Medicaid and CHIP programs coordinate with each other better facilitate transitions for children without causing them to lose coverage (Kaiser, 2012). Collection of demographic data, data pertaining to reasons for disenrollment, and eligibility decisions will enable policymakers to successfully evaluate retention processes (Southern Institute, 2009). The elimination by CHIPRA of the five-year waiting period previously needed for immigrants to receive public insurance is an opportunity to reduce the generational gap of Latino uninsured children (DeCamp, 2012).

Prevalence of Issue

Studies have used the Medical Expenditure Panel Survey, the National Health Interview Survey, and the National Survey of Children's Health to determine the prevalence of uninsurance and unstable coverage among children in the United States. The reported numbers range from 9 to 11.1 million for children with gaps in coverage and 5 to 6 million for those with no insurance in a given survey year (Bethell, 2011; Satchell, 2005). There was a decrease from 10.9% to 10.0% in uninsured children from 2007 to 2010, which is being attributed to children gaining coverage through Medicaid and CHIP as a result of CHIPRA (Kaiser, 2012).

Fiscal Burden of Lack of Insurance and Gaps in Coverage

For low income families, the financial burden is lower for those with full-year public coverage compared to those with full-year private insurance (Galbraith, 2005). Churning establishes additional administrative costs. Although data on the financial impact of churning is limited, Fairbrother (2005) estimated that in California alone, the cost per beneficiary of

re-enrolling in Medi-Cal and a subsequent managed care plan is $180, summing to a total of $140 million dollars per year to re-enroll eligible children who had dropped coverage within a three-year time period. For the financial burden of a community, a 10% disenrollment would increase the costs of health care by $3,460,398 annually, or $2,121 for each disenrolled child as ED visits and hospital stays would increase (Rimsza, 2007). A study of Massachusetts residents that use behavioral health services found that MassHealth closes approximately 34,000 cases per month of which 11,000 are reopened within 90 days at an estimated cost per case for reopening of $200 (Capoccia, 2013).

Child Health Outcomes

Continuity of insurance coverage has the potential to impact child and adolescent health in a number of ways. First, continuous coverage without gaps can permit children and adolescents access to a regular source of care and therefore reduce unmet needs (Aiken, 2004; Holl, 2000; Schoen, 2000). A regular source of care allows for treatment of chronic health conditions, provides routine preventive care, and management of acute and urgent problems (Olson, 2005). The likelihood of receiving preventative care is increased when a child has both a usual source of care and is insured (DeVoe, 2012). Second, continuous coverage ensures that children and adolescents can receive continuity of care without gaps. Continuity of care helps maintain information exchange, coordination of management plans, and ongoing relationships between patients and clinicians (Haggerty, 2003). Continuity of care also permits children's health conditions to be monitored regularly and treatments adjusted to maximize health and prevent exacerbations or worsening of conditions that might lead to hospitalization (Fairbrother, 2004; Weissman, 1992). Third, continuity of coverage may allow time for greater engagement with clinicians in treatment decisions that lead to greater satisfaction with services and better health status (Holl, 2000; Kenney, 2007; Shone, 2005).

With continuity and gaps in coverage having been far less studied than uninsurance status at a point in time, there is limited literature relating continuity with child health outcomes. A cross-sectional analysis of the NSCH 2003 to 2004 described children with continuous private coverage as being the least likely to report having poor or fair health and also least likely to describe their asthma severity as 'minor' when compared to children with continuous public coverage, those who experienced gaps, or the continuously uninsured (Halterman, 2008). An analysis of the National Health and Nutrition Examination Survey found that being insured increased the likelihood that children with intermittent asthma would receive a diagnosis and subsequent control medication (Coker, 2012). Olson et al. (2005) found that more children with public coverage (4.6%), part-year uninsured (2.6%), and full-year uninsured (2.2%) self-reported to have fair or poor health when compared to children with full year private coverage (0.9%). A study looking at children in Georgia found that children who were never/intermittently insured were less likely to view their care as high/moderate quality (Ogbuanu, 2012b).

Policy Factors

From a global perspective, policy changes affect insurance coverage patterns and health outcomes of children. As described previously, many states have implemented various strategies for streamlining and simplifying the enrollment and renewal processes for Medicaid and CHIP. In a California-based study examining two cohorts of children before and after a policy change extending the Medicaid eligibility redetermination period from 3 to 12 months, Bindman et al. (2008b) found that more children had continuous Medicaid coverage and a reduction in hospitalization for ambulatory care sensitive conditions post policy change. In another California-based cross-sectional study Millet et al. (2010) demonstrated that individuals in counties with a choice of Medicaid plans were less likely to have continuous enrollment and higher annual ambulatory care sensitive admission rates than individuals in counties with no choice of Medicaid plans. Changes in children's uninsurance rates from 2007 to 2010 were attributed to policy changes by multiple states (Kaiser, 2012). A 2013 study showed that the seven states which adopted a continuous eligibility policy in 2009 were able to increase average length of child enrollment (Ku, 2013). Generating additional evidence that elucidates the pathway that encompasses policy context, insurance coverage, service delivery, and outcome is critical.

1. Addiss, D. G., N. Shaffer, B. S. Fowler and R. V. Tauxe. 1990. "The Epidemiology of Appendicitis and Appendectomy in the United States." American Journal of Epidemiology 132 (5): 910-25.
2. Aiken, K.D., Freed, G.L., Davis, M.M. (2004). When insurance status is not static: insurance transitions of low-income children and implications for health and health care. Ambul Pediatr, 4(3):237-43.

3. Angier, H., DeVoe, J. E., Tillotson, C., Wallace, L., & Gold, R. (2013). Trends in Health Insurance Status of US Children and their Parents, 1998–2008. Maternal and child health journal, 17(9), 1550-1558.
4. Banerjee, R., Ziegenfuss, J. Y., & Shah, N. D. (2010). Impact of discontinuity in health insurance on resource utilization. BMC Health Services Research, 10, 195.
5. Berdahl, T. A., Friedman, B. S., McCormick, M. C., & Simpson, L. (2013). Annual report on health care for children and youth in the United States: trends in racial/ethnic, income, and insurance disparities over time, 2002–2009.Academic pediatrics, 13(3), 191-203.
6. Bethell, C.D., Kogan, M.D., Strickland, B.B., Schor, E.L., Robertson, J., Newacheck, P.W. (2011) A national and state profile of leading health problems and health care quality for US children: Key insurance disparities and across-state variations. Acad Ped, 11:S22-S33.
7. Bindman, A. B., Chattopadhyay, A., & Auerback, G. M. (2008a). Interruptions in Medicaid coverage and risk for hospitalization for ambulatory care-sensitive conditions. Annals of Internal Medicine, 149(12), 854-860.
8. Bindman, A. B., Chattopadhyay, A., & Auerback, G. M. (2008b). Medicaid re-enrollment policies and children's risk of hospitalizations for ambulatory care sensitive conditions. Medical Care, 46(10), 1049-1054.
9. Byrd, V. L. H. and J. Verdier. 2011. (October). "Collecting, Using, and Reporting Medicaid Encounter Data: A Primer for States. Final Report. October 19, 2011" [accessed on March 27, 2014]. Available at: http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/downloads/MAX_PDQ_Task_X_EncounterDataPrimerforStates.pdf
10. Callahan & S.T., Cooper, W.O. (2007). Continuity of health insurance coverage among young adults with disabilities. Pediatrics, 119(6): 1175-1180.
11. Capoccia, V., Croze, C., Cohen, M., & O'Brien, J. P. (2013). Sustaining enrollment in health insurance for vulnerable populations: Lessons from Massachusetts. Psychiatric Services, 64(4), 360-365.
12. Cassedy, A., Fairbrother, G., & Newacheck, P. W. (2008). The impact of insurance instability on children's access, utilization, and satisfaction with health care. Ambulatory Pediatrics, 8(5), 321-328.
13. Coburn, A. F., McBride, T. D., & Ziller, E. C. (2002). Patterns of health insurance coverage among rural and urban children. Medical Care Research & Review, 59(3), 272-292.
14. Coker, T. R., Kaplan, R. M., & Chung, P. J. (2012). The association of health insurance and disease impairment with reported asthma prevalence in US children. Health services research, 47(1pt2), 431-445.
15. Cummings, J. R., Lavarreda, S. A., Rice, T., & Brown, E. R. (2009). The effects of varying periods of uninsurance on children's access to health care. Pediatrics, 123(3), e411-8.
16. Daly R. Now entering crisis mode. Medicaid woes have states seeking drastic solutions, as feds struggle to keep program strong for patients. Mod Healthc. 2011;41(9):6-7, 16, 11
17. DeCamp, L. R., & Bundy, D. G. (2012). Generational status, health insurance, and public benefit participation among low-income Latino children. Maternal and child health journal, 16(3), 735-743.
18. DeLone, S.E. & Hess, C.A. (2011). Medicaid and CHIP children's healthcare quality measures: What states use and what they want. Acadamic Ped, 11:S69-S76.
19. DeVoe, J. E., Tillotson, C. J., Wallace, L. S., Lesko, S. E., & Pandhi, N. (2012). Is health insurance enough? A usual source of care may be more important to ensure a child receives preventive health counseling. Maternal and child health journal, 16(2), 306-315.
20. DeVoe, J. E., Graham, A., Krois, L., Smith, J., & Fairbrother, G. L. (2008). "Mind the gap" in children's health insurance coverage: Does the length of a child's coverage gap matter? Ambulatory Pediatrics, 8(2), 129-134.
21. Fairbrother, G. (2005). How much does churning in Medi-Cal cost? Los Angeles, CA: The California Endowment and The Common Wealth Fund.
22. Fairbrother, G., Jain, A., Park, H. L., Massoudi, M. S., Haidery, A., & Gray, B. H. (2004). Churning in Medicaid managed care and its effect on accountability. Journal of Health Care for the Poor & Underserved, 15(1), 30-41.
23. Fairbrother G, Madhavan G, Goudie A, et al. Reporting on continuity of coverage for children in Medicaid and CHIP: What states can learn from monitoring continuity and duration of coverage. Acad Pediatr. 2011;11(4):318-325
24. Federico, S. G., Steiner, J. F., Beaty, B., Crane, L., & Kempe, A. (2007). Disruptions in insurance coverage: Patterns and relationship to health care access, unmet need, and utilization before enrollment in the state children's health insurance program. Pediatrics. 120(4), e1009-16.

25. Feudtner C., Christakis D.A., & Connell F.A. (2000). Pediatric Deaths Attributable to Complex Chronic Conditions: A Population-Based Study of Washington State, 1980–1997. Pediatrics. 106(1):205-209.
26. Feudtner C., Hays R.M., Haynes G., Geyer J.R., Neff J.M., & Koepsell T.D. (2001). Deaths Attributed to Pediatric Complex Chronic Conditions: National Trends and Implications for Supportive Care Services. Pediatrics. 107(6):e99.
27. Flores, G., & Lin, H. (2013). Trends in racial/ethnic disparities in medical and oral health, access to care, and use of services in US children: has anything changed over the years?. Int J Equity Health, 12(10).
28. Flores, G. & Tomany-Korman, S.C. (2008). Racial and ethnic disparities in medical and dental health, access to care, and use of services in US children. Pediatrics. 121 (2): E286-E298 .
29. Fowler F.J., Gallagher P.M., & Homer C.J. (2001). Identifying children with special needs. Seventh Conference on Health Survey Research Methods. Department of Health and Human Services. Maryland.
30. Galbraith, A.A., Wong, S.T., Kim, S.E., Newacheck, P.W. (2005). Out-of-pocket financial burden for low-income families with children: socioeconomic disparities and effects of insurance. Health Services Research, 40(6 Pt 1):1722-36.
31. General Services Administration. CHIP Annual Report Template System (CARTs). Washington, DC: 2011. Available at: http://www.reginfo.gov/public/do/PRAViewIC?ref_nbr=201111-0938-009&icID=200547. Accessed May 6, 2013
32. Ginde, A. A., Lowe, R. A., & Wiler, J. L. (2012). Health insurance status change and emergency department use among US adults. Archives of internal medicine, 172(8), 642-647.
33. Haggerty, J.L., Reid, R.J., Freeman, G.K., Starfield, B.H., Adair, C.E., McKendry, R. (2003). Continuity of care: a multidisciplinary review. BMJ, 327(7425):1219-21.
34. Halterman J.S., Montes G., Shone L.P., & Szilagyi P.G. (2008). The impact of health insurance gaps on access to care among children with asthma in the United States. Ambulatory Pediatrics. 8:43-49.
35. Hauptman, J. S., Dadour, A., Oh, T., Baca, C. B., Vickrey, B. G., Vassar, S., ... & Mathern, G. W. (2013). Time to Pediatric Epilepsy Surgery Is Longer and Developmental Outcomes Lower for Government Compared With Private Insurance. Neurosurgery, 73(1), 152-157.
36. Holl, J. L., Szilagyi, P. G., Rodewald, L. E., Shone, L. P., Zwanziger, J., Mukamel, D. B., et al. (2000). Evaluation of New York State's Child Health Plus: Access, utilization, quality of health care, and health status. Pediatrics, 105(3 Suppl E), 711-718.
37. Ireys H.T., Anderson G.F., Shaffer T.J., Neff J.M. (1997). Expenditures for care of children with chronic illnesses enrolled in the Washington State Medicaid program, fiscal year 1993. Pediatrics. 100(2 Pt 1):197-204.
38. Kaiser Commission on Medicaid and the Uninsured (2012) Secrets to Success: An Analysis of Four States at the Forefront of the Nation's Gains in Children's Health Coverage. The Henry J. Kaiser Family Foundation. Washington, DC.
39. Kenney, G. (2007). The impacts of the State Children's Health Insurance Program on children who enroll: findings from ten states. Health Services Research, 42(4):1520-43.
40. Kenney GM, Cook A, Dubay L. Progress enrolling children in Medicaid/CHIP: Who is left and what are the prospects for covering more children? Washington, DC: Urban Institute; December 2009. Available at: http://www.urban.org/research/publication/progress-enrolling-children-medicaidchip-who-left-and-what-are-prospects-covering-more-children-updated-121409. Accessed January 18, 2009
41. Kogan, M. D., Newacheck, P. W., Blumberg, S. J., Ghandour, R. M., Singh, G. K., Strickland, B. B., et al. (2010). Underinsurance among children in the United States. New England Journal of Medicine, 363(9), 841-851.
42. Kreiger N., Williams D.R., & Moss N.E. (1997). Measuring Social Class in US Public Health Research: Concepts, Methodologies, and Guidelines. Annu. Rev. Public Health. 18:341-78.
43. Ku, L., P. MacTaggart, F. Pervez and S. Rosenbaum. 2009. (July). "Improving Medicaid's Continuity of Coverage and Quality of Care" [accessed on July 26, 2012]. Available at: http://www.communityplans.net/Portals/0/ACAP%20Docs/Improving%20Medicaid%20Final%20070209.pdf.
44. Ku, L., E. Steinmetz and B. K. Bruen. 2013. "Continuous-Eligibility Policies Stabilize Medicaid Coverage for Children and Could Be Extended to Adults with Similar Results." Health Affairs 32 (9): 1576-82.
45. Levinson, D. R. 2009. (May). "Medicaid Managed Care Encournter Data: Collection and Use" [accessed on March 27, 2014]. Available at: http://oig.hhs.gov/oei/reports/oei-07-06-00540.pdf.

46. Millett, C., Chattopadhyay, A., & Bindman, A. B. (2010). Unhealthy competition: Consequences of health plan choice in California Medicaid. American Journal of Public Health, 100(11): 2235-2240.
47. National Conference of State Legislatures Health Policy Tracking Service. The state Medicaid crisis: Benefits and eligibility on the block. Healthc Financ Manage. 2003;57(5):14, 16
48. Neuzil K.M., Wright P.F., Mitchel E.F. Jr, Griffin M.R. (2000). The burden of influenza illness in children with asthma and other chronic medical conditions. J Pediatr. 137(6):856-64.
49. Ogbuanu, C., Goodman, D., Kahn, K., Noggle, B., Long, C., Bagchi, S., ... & Castrucci, B. (2012a). Factors associated with parent report of access to care and the quality of care received by children 4 to 17 years of age in Georgia. Maternal and child health journal, 16(1), 129-142.
50. Ogbuanu, C., Goodman, D. A., Kahn, K., Long, C., Noggle, B., Bagchi, S., ... & Castrucci, B. (2012b). Timely access to quality health care among Georgia children ages 4 to 17 years. Maternal and child health journal, 16(2), 307-319.
51. Okumura, M. J., McPheeters, M. L., & Davis, M. M. (2007). State and national estimates of insurance coverage and health care utilization for adolescents with chronic conditions from the national survey of children's health, 2003. Journal of Adolescent Health, 41(4), 343-349.
52. Olson, L. M., Tang, S. F., & Newacheck, P. W. (2005). Children in the United States with discontinuous health insurance coverage. New England Journal of Medicine, 353(4), 382-391.
53. Park, J. H. G., Linakis, J. G., Skipper, B. J., & Scott, S. M. (2012). Factors that Predict Frequency of Emergency Department Utilization in Children With Diabetes-Related Complaints. Pediatric emergency care, 28(7), 614-619.
54. Patrick SW, Choi H, Davis MM. Increase in federal match associated with significant gains in coverage for children through Medicaid and CHIP. Health Aff (Millwood). 2012;31(8):1796-1802
55. Percheski, C., & Bzostek, S. (2013). Health insurance coverage within sibships: Prevalence of mixed coverage and associations with health care utilization. Social Science & Medicine, 90, 1-10.
56. Rimsza, M.E., Butler, R.J., Johnson, W.G. (2007). Impact of Medicaid Disenrollment on Health Care Use and Cost. Pediatrics., 119:pp e1026-e1032.
57. Rosenbaum P.R. (2010). Design of Observational Studies. (pp 43-44; 172-175; 188; 197-206). New York, NY: Springer.
58. Sabik, L. M., & Dahman, B. A. (2012). Trends in care for uninsured adults and disparities in care by insurance status. Medical Care Research and Review,69(2), 215-230.
59. Satchell, M. & Pati, S. (2005). Insurance gaps among vulnerable children in the United States, 1999-2001. Pediatrics, 116(5):1155-61.
60. Schoen, C. & DesRoches, C. (2000). Uninsured and unstably insured: the importance of continuous insurance coverage. Health Serv Res,35(1 Pt 2):187-206.
61. Seferian E.G., Lackore K.A., Rahman A.S., Naessens J.M., Williams A.R. (2006). Comparison of chronic illness among children receiving mechanical ventilation in a cohort of children's hospitals in 1991 and 2001. J. Peds. 149(6):788-792.e2.
62. Shone, L.P., Dick, A.W., Klein, J.D., Zwanziger, J., Szilagyi, P.G. (2005). Reduction in racial and ethnic disparities after enrollment in the State Children's Health Insurance Program. Pediatrics, 115(6):e697-705.
63. The Southern Institute on Children and Families (2009). Medicaid and CHIP retention: A key strategy to reducing the uninsured. Columbia, SC: Southern Institute on Children and Families.
64. Tillotson, C. J., Wallace, L. S., Lesko, S. E., & Angier, H. (2012). The effects of health insurance and a usual source of care on a child's receipt of health care. Journal of Pediatric Health Care, 26(5), e25-e35.
65. Todd J., Armon C., Griggs A., Poole S., & Berman S. (2006). Increased rates of morbidity, mortality, and charges for hospitalized children with public or no health insurance as compared with children with private insurance in Colorado and the United States. Pediatrics. 118(2):577-585.
66. Valentine J.M., Neff J., Park A.N., Sharp V., Maynard C., Christakis D., Hicks-Thomson J. (2000). Pediatric Hospitalization Patterns for Selected Chronic Health Conditions Using Hospital Abstract Reporting System Data: Methods and Findings. Health Services and Outcomes Research Methodology 1:3-4:335-350.
67. Weissman, J.S., Gatsonis, C., Epstein, A.M. (1992). Rates of avoidable hospitalization by insurance status in Massachusetts and Maryland. JAMA. 1992,268(17):2388-94.

68. Wellek, S. 2010. Testing Statistical Hypotheses of Equivalence and Noninferiority. Florida: Chapman & Hall/CRC, 2nd Edition.
69. Yu, J., Harman, J. S., Hall, A. G., & Duncan, R. P. (2011). Impact of Medicaid/ SCHIP disenrollment on health care utilization and expenditures among children: A longitudinal analysis. Medical Care Research & Review, 68(1), 56-74.
70. Ziller, E. C., Lenardson, J. D., & Coburn, A. F. (2012). Health Care Access and Use Among the Rural Uninsured. Journal of health care for the poor and underserved, 23(3), 1327-1345.

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	

Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

A literature review was performed to identify literature related to effects of uninsurance and gaps in coverage, specifically in the U.S.

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

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

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

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

Evidence_Silber_3154_v6.docx

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

Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

1b. Performance Gap

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

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

1b.1. **Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*
IF a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)
IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.

States are frequently asked to determine public insurance participation rates or measure continuity of enrollment among vulnerable children, both for federal compliance audits and performance-based incentives, and for internal studies concerning vulnerable populations (Patrick et al., 2012; Daly 2003; National Conference of State Legislatures Health Policy Tracking Service, 2003). Participation rates are defined as the fraction of eligible children who are enrolled (Kenney et al., 2009). We developed and validated this administrative claims–based participation metric, "Informed Coverage," using a naturally occurring randomization observed inside each state that dynamically informs assumptions about patterns of eligibility and allows statewide estimates of participation rates using only administrative claims data. This standardized measure can be used by states as a potential indicator of quality and access. The issue of enrollment and retention is a long-standing concern for publically financed insurance programs, and one that states have likely examined using less formal means. Because Medicaid/CHIP enrollees are from low-income families, this measure will benefit vulnerable children it will hold states accountable for retaining children eligible for public coverage. Where data capacity permits, this measure also takes into account children switching from Medicaid and CHIP and vice-versa instead of treating children as disenrolled from public insurance.

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

The Informed Coverage fraction distributions across states for the time period January 1, 2008 to June 30, 2009 are presented in Table 1 of the Appendix, as are the distributions for the intermediate calculations. The mean state Informed Coverage value was 0.7949 (SD 0.1035). The minimum state Informed Coverage Value was 0.3814, the maximum was 0.9350, and the IQR was 0.7474, 0.8580. The deciles of the state Informed Coverage values from 10% through 90% were as follows: 0.6623, 0.7326, 0.7576, 0.7778, 0.8261, 0.8469, 0.8571, 0.8618, 0.8840. Table 2 of the Appendix provides the number and percentage of children included in the Coverage PE and Coverage PI intermediate calculations (and thus Informed Coverage) by state out of all children present in the MAX dataset for the years 2008 to 2009. Data quality limited analyses to 43 states. Six states failed to have sufficient reporting of managed care claims for the data utilized in the development of this measure and thus were eliminated from analyses for this time period and dataset: Kentucky, Massachusetts, Mississippi, Ohio, Pennsylvania, and West Virginia. Additionally, Maine and the District of Columbia were found to have excessive quality issues in their inpatient records and were

likewise eliminated. Table 3 in the Appendix provides all state's intermediate calculations (Coverage PE, Coverage PI, Appendectomy Coverage Rate) and final Informed Coverage value.

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

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*
Disparities in continuity of coverage according to ethnicity, geography, insurance type, and special health care need have been observed throughout the literature. Publicly insured children from poorer households are more likely than those from higher income households to have gaps in insurance coverage (Bethell 2011; Angier 2013). Minority children, especially Hispanic children, are more likely to be uninsured, have gaps in coverage, and not have an usual source of care (Federico 2007; Flores 2008; Kogan 2010; Flores 2013; Berdahl 2013). This ethnic disparity is worse for first- and second-generation Latino children (DeCamp 2012). Minorities in Georgia were also found to have lower access to higher quality of healthcare (Ogbuanu 2012a). Rural children are more likely to have longer periods of time without insurance than children in urban settings (Coburn 2002). There is a larger gap between insured and uninsured for children in urban settings than in rural settings, and children in urban settings are less likely to have a usual source of care regardless of insurance status (Ziller 2012). Children with special health care needs are more likely to have public insurance coverage than private insurance coverage but experience more unmet needs (Bethell 2011; Callahan 2007; Okumura 2007). Olson, Tang, and Newacheck (2005) performed a cross-sectional study using National Health Interview Surveys confirming that children with full-year public insurance coverage report a higher prevalence of chronic conditions limiting activities relative to children with full year private insurance coverage (12.3% vs 5.1%). In specific patient populations, such as those with diabetes related complaints, the insurance of children with diabetes was Medicaid rather than private (Park 2012). Children with public insurance had a longer interval between epilepsy seizure onset and referral and subsequent surgery compared to privately insured children (Hauptman 2013).

Race/ethnicity (Appendix Table 4): For these analyses, race and ethnicity was determined based on the race-ethnicity variable reported in the MAX data and classified based on Office of Management and Budget guidelines. White was defined as White, not of Hispanic origin. Black was defined as Black, not of Hispanic origin. For Hispanic, we combined children reported as 'Hispanic or Latino' and 'Hispanic or Latino and one or more races'. Other included American Indian, Alaskan Native, Asian, Pacific Islander and children with missing race/ethnicity. We stratified the 18-month informed coverage fraction by enrollee race/ethnicity. Coverage fractions varied by race within and across states with variations in the race/ethnicity groups with the highest and lowest Informed Coverage values within the state.

Special Health Care Needs (Appendix Table 5): Based on published peer-reviewed literature, we compiled a list of pediatric chronic conditions where each condition was represented in all or most of the papers (Valentine, 2000; Ireys, 1997; Todd, 2006; Fowler, 2001; Neuzil, 2000; Feudtner, 2000; Feudtner, 2001; Seferian, 2006). We stratified the 18-month informed coverage fraction by enrollee chronic condition status. Informed Coverage value were generally higher for children with chronic care needs.

Socioeconomic Status (Appendix Table 6-8): Socioeconomic measures at the individual or census-tract level are not included in the MAX data. Although five digit-zip code-based socioeconomic measures have significant limitations, we performed analyses using three socioeconomic variables (% with high school degree, % with income below federal poverty level, and income level) stratified by quartiles in order to demonstrate that these analyses are feasible (Krieger, 1997). These variables were abstracted from U.S. census 5-digit-zip code-level data and merged with the data. If 9-digit-zip code data were available in the MAX data, these analyses would produce more robust and meaningful results. As noted in the methods, these analyses were performed for the purposes of demonstrating feasibility and not for the purposes of assessing the significance of associations. Informed Coverage did not vary by much between the highest and lowest quartiles for any measure. The coverage fraction across the poverty quartiles lacked a coherent pattern, although the extremes show that ZIP codes with a lower percentage of enrollees below the FPL had better coverage than ZIP codes with a higher percentage above the FPL. Differences across income quartiles were also small and lacked a coherent pattern across states. When looking at just the extremes the lowest income quartile always

had a better coverage than those in the highest quartile. Trends for education were clear: while coverage fractions were generally homogenous and never differed by more than ten percentage points between the most- and least-education quartiles, coverage fractions in every state improved as the high school graduation rate of enrollee ZIP codes fell.

Rurality/Urbanity (Appendix Table 9): A crosswalk was performed between the MAX data using the 2010 Census urban and rural classification (http://www.census.gov/geo/www/ua/2010urbanruralclass.html). There are two types of urban areas: urbanized areas have 50,000 or more people residing in that area; urban clusters have at least 2,500 and less than 50,000 people residing in that area. Rural area encompasses all population, housing, and territory not included within an urban area. The Informed Coverage values were drastically lower for enrollees who lacked a geographic status (zip code information missing), compared to any other category. Informed Coverage values were generally similar between urban, rural, and urban cluster areas in each state.

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

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

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

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://cor.research.chop.edu/node/26

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

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
No data dictionary **Attachment:**

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

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

N/A

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator for Informed Coverage represents the sum (within a state) of months enrolled in Medicaid/CHIP for all children over an 18-month window.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator is the summation (within a state) of months enrolled in Medicaid/CHIP for all children (0-18 years) over an 18-month window. A month is considered "covered" if a child has greater than 14 enrolled days in that month or if there is an indicator for S-CHIP coverage for that month. Figures 1 and 2 in the Appendix provide an illustration of Coverage PE and Coverage PI.

To determine what is the best assumption to use (either the Appendectomy Coverage Rate (or ACR), PI, or PE) inside each state, we compare the observed appendectomy coverage rate in a state, to the estimated coverage rate that would be calculated in that state with either PI, or PE assumptions. If PE < ACR < PI, we utilize ACR. If ACR > PI, we use PI, and if ACR < PE we use PE.

The ACR reflects a natural experiment since appendicitis is a random event, not dependent on healthcare of SES status. Appendicitis is defined using principal diagnosis (ICD-9 CM codes 540-541 Appendicitis; ICD-10 codes K35.2, K35.3, K35.80, K35.89, K37) or procedure (ICD-9 CM 47.0-47.09, 47.2 Appendectomy; ICD-10 codes 0DTJ4ZZ, 0DTJ0ZZ, 0DTJ7ZZ, 0DTJ8ZZ, 0D9J00Z, 0D9J0ZZ, 0D9J30Z, 0D9J3ZZ, 0D9J40Z, 0D9J4ZZ, 0D9J70Z, 0D9J7ZZ, 0D9J80Z, 0D9J8ZZ). This condition is utilized as it (1) has an acute onset (reflecting a discrete point in time); (2) has an incidence rate that is not influenced by prior care, insurance coverage, or by factors that may influence obtaining coverage, such as socioeconomic status; and, (3) would require hospitalization for all children regardless of insurance status. If a child is hospitalized and generates a bill seen in the Medicaid claims, they must have been eligible for Medicaid. If a child was not enrolled at the time of developing appendicitis, but was eligible, the appendicitis should still be observed because Medicaid and most CHIP programs allow up to three months of retroactive coverage and most states have policies of presumptive eligibility for their public insurance program. By identifying appendicitis hospitalizations and determining whether these children were enrolled prior to their hospitalization, we can utilize the rate of existing enrollment at the specific time point of the event to estimate the participation rate for the state population (number enrolled over number eligible at a given point in time). We determine if a child was enrolled prior to hospitalization using a look-back to their state of enrollment 4 months prior to hospitalization. The numerator for the appendicitis calculation is the number of children with an appendicitis hospitalization during the same 18-month observation window used for the Coverage PE and Coverage PI intermediate calculations, who are enrolled in Medicaid/CHIP four months prior to their inpatient stay.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
TThe sum (within a state) of months eligible for Medicaid/CHIP for all children (0-18 years) over an 18-month window. In addition, months that could be defined as "eligible" are based on known events recorded in the MAX data that would affect eligibility (birth or ageing out).

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
For the intermediate calculations of "Coverage Presumed Eligible (PE)" and "Coverage Presumed Ineligible (PI)", the denominator is the summation (within a state) of the months a child is eligible for Medicaid/CHIP over an 18-month observation window. The assumptions used to define a child as "eligible" for Medicaid/CHIP coverage for a given month is specific to which intermediate

computation is being calculated. When calculating the intermediate computation of "Coverage Presumed Eligible (PE)", a child is defined as being eligible based on an 18-month observation, in combination with an 18-month look-back period. If any enrollment is observed in the 18-month look-back period, the child is defined as eligible for the entire 18-month observation window. If there is no evidence of enrollment in the 18-month look-back period, eligibility is defined from the first point of enrollment in the observation window. When calculating the intermediate computation of "Coverage Presumed Ineligible (PI)", a child is defined as being eligible solely on the 18-month observation window. For Coverage PI, eligibility starts from the first enrolled month during the 18-month observation window.

Again using the point-in-time analysis of appendicitis to calculate the observed participation rate, the denominator for the appendicitis calculation, is the number of children with an appendicitis hospitalization during the same 18-month observation window used for the Coverage PE and Coverage PI intermediate calculations. Appendicitis is defined using principal diagnosis (ICD-9 CM codes 540-541 Appendicitis; ICD-10 codes K35.2, K35.3, K35.80, K35.89, K37) or procedure (ICD-9 CM 47.0-47.09, 47.2 Appendectomy; ICD-10 codes 0DTJ4ZZ, 0DTJ0ZZ, 0DTJ7ZZ, 0DTJ8ZZ, 0D9J00Z, 0D9J0ZZ, 0D9J30Z, 0D9J3ZZ, 0D9J40Z, 0D9J4ZZ, 0D9J70Z, 0D9J7ZZ, 0D9J80Z, 0D9J8ZZ). Appendicitis was chosen because the aim was to create a population where both enrolled and unenrolled eligible children are identifiable in MAX, we sought a condition that: (1) has an acute onset (reflecting a discrete point in time); (2) has an incidence rate that is not influenced by prior care, insurance coverage, or by factors that may influence obtaining coverage, such as socioeconomic status; and, (3) would require hospitalization for all children, regardless of insurance status. Appendicitis meets these three criteria. Appendicitis has an acute onset which occurs at random and is not influenced by previous care or insurance status; it is not influenced by child or parental characteristics or actions that affect likelihood of coverage; and if children develop appendicitis, they will be hospitalized. If a child is hospitalized and generates a bill seen in the Medicaid claims, they must have been eligible for Medicaid. If a child was not enrolled at the time of developing appendicitis, but was eligible, the appendicitis should still be observed because Medicaid and most CHIP programs allow up to three months of retroactive coverage and most states have policies of presumptive eligibility for their public insurance programs.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
FFor the appendicitis calculation, the population is limited to children between the ages of 2 to 16 years old. To determine what is the best assumption to use (either the Appendectomy Coverage Rate (or ACR), PI, or PE) inside each state, we compare the observed appendectomy coverage rate in a state, to the estimated coverage rate that would be calculated in that state with either PI, or PE assumptions.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
CFor children who are born within the 18-month window of observation, the total months of eligibility begins from date of birth. Finally, for children who reach the age of 18 before the end of the 18-month window of observation, the total month of eligibility ends with their 18th birthday.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*
Stratification can be performed for Informed Coverage using any desired strata that policymakers choose to study. For example, stratification can be performed within states based on the type of Medicaid and CHIP programs, or by race.

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)
No risk adjustment or risk stratification
If other:

S.12. Type of score:
Other (specify):
If other: Informed Coverage is utilized to better determine the participation rates for states reporting MAX data. The eligibility assumptions for each state are determined by comparing the approaches of Presumed Eligible and Presumed Ineligible rates. Higher estimates of participation rates (higher Informed Coverage) reflects better quality of enrollment for programs within each state.

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

Better quality = Higher score

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

The following describes the steps for calculating the intermediate computations and their use for the final determination. A minimum of three continuous years of MAX claims data are required. The first 18 months are used for a lookback and the second 18 months are the observation period. The same 18-month observation window is used for all calculations. All calculations are done within a state.

Determine the appendectomy participation rate (APR) Intermediate Calculation:
The prior participation of eligible patients developing appendicitis 4 months prior to developing appendicitis
Step 1- Calculate the denominator for appendectomy participation rate: 1) Identify all children between the ages 2 and 16 at the start of the 18-month observation window; 2) Identify the number of children with an inpatient admission for either a principal diagnosis of appendicitis (ICD-9 CM codes 540-541; ICD-10 codes K35.2, K35.3, K35.80, K35.89, K37) or a principal procedure of appendectomy (ICD-9 CM codes 47.0-47.09, 47.2; ICD-10 codes 0DTJ4ZZ, 0DTJ0ZZ, 0DTJ7ZZ, 0DTJ8ZZ, 0D9J00Z, 0D9J0ZZ, 0D9J30Z, 0D9J3ZZ, 0D9J40Z, 0D9J4ZZ, 0D9J70Z, 0D9J7ZZ, 0D9J80Z, 0D9J8ZZ). Step 2- calculate the numerator for appendectomy coverage rate: 1) Identify the total number of children with pre-existing enrollment in Medicaid or CHIP. Pre-existing enrollment is defined as an observed enrollment exactly four months prior to their date of admission. Step 3- Calculate the appendectomy participation rate: compute the percentage of children admitted for appendicitis/appendectomy with pre-existing enrollment in Medicaid or CHIP, defined by enrollment 4 months prior to the admission.

Determination of the Appendectomy Never Participated Rate (ANPR) Intermediate Calculation: The fraction of eligible appendectomy patients who did not have any participation noted at any point 4 or more months prior to developing appendicitis (within the limits of the observation and lookback period data).

Coverage PE Intermediate Calculation:
Step 4- To determine the denominator for Coverage PE (total months of eligibility using the PE approach): 1) identify all children enrolled in Medicaid/CHIP at any point within the 18-month window of observation AND/OR the 18-month look back, excluding those older than 18 at the beginning of the 18-month observation window; 2) Identify all children who are born within the 18-month window of observation – for these children, total months of eligibility begin from date of birth; 3) Identify all children who reach the age of 18 before the end of the 18-month window of observation – for these children, total months of eligibility end with their 18th birthday; 4) Identify all children who DO NOT APPEAR as covered at any point within the 18-month look back period (“covered” defined as at least one day of coverage) – for these children, total months of eligibility begin with their first day of coverage within the 18-month observation window; 5) For all other children who do not represent populations in Steps 1, 2, or 3, total months of eligibility equals all 18 months in the observation window; and 6) The Coverage PE denominator is the summation of total number of eligible months for all children in the eligible population. Step 5- to determine the numerator for Coverage PE (total months of coverage using PE approach): 1) Identify total number of months in the 18 month observation window covered by MAX/CHIP for each child in the eligible population. A month is considered “covered” if the child has greater than 14 days of enrollment in that month or if there is an indicator for S-CHIP coverage for that month; and 2) The Coverage PE numerator is the summation of total months covered within the 18-month observation window for all children in the eligible population. Step 6- Calculate the Coverage PE intermediate value: compute the percentage of months covered within the 18-month observation window (Coverage PE numerator divided by Coverage PE denominator).

PE adjustment for patients never enrolled (PE´): See appendix for derivation (Figure 3). PE´=PE*(1-ANPR).

Coverage PI Intermediate Calculation:
Step 7- To determine the denominator for Coverage PI (the total months of eligibility using the PI approach): 1) identify all children enrolled in Medicaid/CHIP at any point within the 18-month window of observation, excluding those children older than 18 at the beginning of the 18-month observation window; 2) Identify all children who are born within the 18-month window of observation – for these children, total months of eligibility begin from date of birth; 3) Identify all children who reach the age of 18 before the end of the 18-month window of observation – for these children, total months of eligibility ends with their 18th birthday; 4) For all other children who do not represent populations in Steps 1, 2, or 3, months of eligibility begins with the first observed enrollment in the observation window and continues for the remainder of the observation window; and 5) The Coverage PI denominator is the summation of the total number of eligible months for all children in the eligible population. Step 8- to determine the numerator for Coverage PI (total months of coverage using PI approach): 1) Identify the total number of

months in the 18-month observation window covered by MAX/CHIP for each child in the eligible population. A month is considered "covered" if the child has greater than 14 days of enrollment in that month or if there is an indicator for S-CHIP coverage for that month; and 2) The Coverage PI numerator is the summation of the total months covered within the 18-month observation window for all children in the eligible population. Step 9- Calculate the Coverage PI intermediate value: compute the percentage of months covered within the 18-month observation window (Coverage PI numerator divided by Coverage PI denominator).

Informed Coverage:
Step 10- The Informed Coverage is the weighted mean of the state Coverage PE′ and state Coverage PI values, where the weights are determined by the state appendectomy participation rate. The closer the appendectomy rate is to Coverage PE, the more weight that Coverage PE receives in the informed coverage measure, and the closer the appendectomy rate is to Coverage PI, the more weight that Coverage PI receives in the informed coverage. An illustration of the formula for this calculation is provided in Figure 4 of the Appendix.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
N/A

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
N/A

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Claims (Only)

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The Medicaid Analytic eXtract (MAX) claims data are used for this metric.

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

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

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
No Applicable Care Setting
If other:

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

2. Validity – See attached Measure Testing Submission Form
Testing_Silber_3154_RR_v3.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

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

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

What were the statistical results of the analyses used to select risk factors?

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

No - This measure is not risk-adjusted

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

Measure Number (*if previously endorsed*)**:** 3154
Measure Title: Informed Coverage
Date of Submission: 12/12/2016
Type of Measure:

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

Instructions

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

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

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

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

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

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

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

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

OR

- rationale/data support no risk adjustment/ stratification.

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

OR

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

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

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

Notes

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

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have

differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

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

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

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

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

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
The Medicaid claims data, Medicaid analytic eXtract (MAX), were utilized for testing both the reliability and the validity of the measure. MAX is a database that contains state enrollment and claims data for children enrolled in Medicaid and the Children's Health Insurance Program (CHIP). These data are collected as part of each state's Medicaid Management Information System (MMIS), which is unique to the state's Medicaid program. To allow for federal monitoring of the

Medicaid program at the national level, the MMIS data are transformed to a uniform database and submitted to Center for Medicare & Medicaid Services (CMS) via the Medicaid and CHIP Statistical Information System (MSIS). Alternatively, the Informed Coverage measure can be used with other data sets that include patient level data of monthly Medicaid/CHIP enrollment and inpatient claims with diagnosis information. The American Community Survey (ACS) for 2008-2010 was used for testing validity.

1.3. What are the dates of the data used in testing? January 2008 – December 2010

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

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

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)
MAX data from a total of 43 states from the United States of America were included in the testing and analysis of this measure. Six states failed to report adequate data for the data used in the testing and analysis, and were not included: Kentucky, Massachusetts, Mississippi, Ohio, Pennsylvania, and West Virginia. Additionally, Maine and the District of Columbia were found to have excessive quality issues in their inpatient records and were likewise excluded.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)
Children enrolled in Medicaid/CHIP in one of the 43 states analyzed during the observation window of January 2008 – June 2009 or July 2009 – December 2010 were included, though those that were 18 at the beginning of the time period were excluded. For the first observation window, 45 million patients were identified with a mean age of 6.98 years of which 49.4% were female, 35.7% were white non-Hispanic, 21.2% were black non-Hispanic, 29.2% Hispanic or Latino, and 13.8% were other. The ACS is conducted by the U.S. census with over 2.8 million households interviewed annually.

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

Two time periods of MAX data were used in testing: 1) a development time period (January 2008 – June 2009); and 2) a validation time period (July 2009 – December 2010). Data from 2006 – 2007 were used as part of the lookback for the development time period data. In addition, the ACS for 2008 – 2010 were used for testing the validity of the MAX findings. Because of its ability to observe eligible unenrolled children, we utilize the ACS survey as an independent data source capable of estimating participation rates in order to validate the new Informed Coverage metric's estimate of

participation. A random split-sample for the 2009 ACS was used with one half being utilized to test the development data and the other half for the validation data.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).
Proxy variables were available and used: sociodemographic variables were determined at the level of a 5-digit zip code: average income, percentage below the Federal Poverty Line, and percentage with high school degree.

2a2. RELIABILITY TESTING

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

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)
State Informed Coverage values for January 2008 – June 2009 were calculated using the weighted mean of the state's Coverage PE and Coverage PI with the weights based on the intermediate calculation's closeness to the state Appendectomy coverage rate. The endpoints (lower, upper) of the 95% Confidence Interval for Informed Coverage are given by the formulas below. Note these formulas provided confidence intervals that were very close to the actual bootstrap estimates (see appendix).

$$lower\ 95\%\ CI\ endpoint = \min\left(\max(lower\ PE', lower\ ACR), lower\ PI\right)$$

$$upper\ 95\%\ CI\ endpoint = \max(\min(upper\ PI, upper\ ACR), upper\ PE')$$

Reliability was assessed by determining if states could be identified with non-overlapping 95% CI.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)
The below table provides the Informed Coverage values with 95% CI for each state for January 2008-June 2009, Figure 1 illustrates these results. Table 1 also includes the Informed Coverage values with 95% CI for each state for the validation time period, Figure 2 illustrates the results for the validation time period.

Table 1. Informed Coverage (95% CI) by State

	Development Time Period	Validation Time Period
AK	0.8257 (0.7545,0.8372)	0.8512 (0.7878,0.9146)
AL	0.8816 (0.8505,0.8922)	0.9287 (0.9060,0.9515)

	Development Time Period	Validation Time Period
AR	0.8694 (0.8435,0.8704)	0.8949 (0.8637,0.9261)
AZ	0.7612 (0.7391,0.7834)	0.8964 (0.8808,0.9080)
CA	0.7993 (0.7889,0.8097)	0.8430 (0.8341,0.8519)
CO	0.6623 (0.6092,0.7154)	0.7342 (0.6889,0.7796)
CT	0.7990 (0.7440,0.8540)	0.8986 (0.8695,0.9276)
DE	0.7778 (0.6174,0.8691)	0.8333 (0.6612,0.9264)
FL	0.6180 (0.5928,0.6431)	0.7248 (0.7049,0.7446)
GA	0.7474 (0.7088,0.7860)	0.7638 (0.7292,0.7984)
HI	0.8571 (0.7412,0.9293)	0.9000 (0.8070,0.9646)
IA	0.8720 (0.8208,0.8846)	0.8454 (0.7945,0.8962)
ID	0.7704 (0.7115,0.8293)	0.8411 (0.7921,0.8901)
IL	0.9350 (0.9332,0.9354)	0.9535 (0.9435,0.9636)
IN	0.8401 (0.8014,0.8788)	0.8747 (0.8408,0.9085)
KS	0.7326 (0.6664,0.7987)	0.7500 (0.6439,0.8561)
LA	0.9235 (0.9020,0.9403)	0.9410 (0.9237,0.9583)
MD	0.8306 (0.7918,0.8693)	0.8694 (0.8396,0.8992)
MI	0.6994 (0.6488,0.7499)	0.8507 (0.8227,0.8787)
MN	0.7231 (0.6776,0.7686)	0.8042 (0.7615,0.8469)
MO	0.8504 (0.8200,0.8808)	0.8860 (0.8593,0.9127)

	Development Time Period	Validation Time Period
MT	0.7083 (0.6033,0.8133)	0.7407 (0.6510,0.8362)
NC	0.7641 (0.7356,0.7926)	0.8022 (0.7762,0.8282)
ND	0.7391 (0.6122,0.8549)	0.9107 (0.8353,0.9131)
NE	0.8333 (0.7649,0.8731)	0.8400 (0.7813,0.8987)
NH	0.8710 (0.7875,0.8919)	0.9167 (0.8467,0.9292)
NJ	0.8533 (0.8284,0.8782)	0.8738 (0.8512,0.8964)
NM	0.8580 (0.8282,0.8877)	0.9065 (0.8832,0.9298)
NV	0.3814 (0.2848,0.4781)	0.4167 (0.3442,0.4891)
NY	0.8884 (0.8792,0.8888)	0.9007 (0.8884,0.9130)
OK	0.8603 (0.8331,0.8612)	0.8584 (0.8325,0.8842)
OR	0.7507 (0.7046,0.7969)	0.7850 (0.7487,0.8213)
RI	0.8618 (0.8008,0.8768)	0.9181 (0.8997,0.9197)
SC	0.7576 (0.7023,0.8128)	0.7339 (0.6882,0.7795)
SD	0.8548 (0.7672,0.8756)	0.9048 (0.8323,0.9254)
TN	0.8571 (0.8225,0.8918)	0.8809 (0.8516,0.9101)
TX	0.6868 (0.6731,0.7004)	0.7257 (0.7130,0.7383)
UT	0.5789 (0.5049,0.6529)	0.6516 (0.5766,0.7266)
VA	0.8225 (0.7828,0.8623)	0.8290 (0.7915,0.8666)
VT	0.9216 (0.8931,0.9235)	0.9508 (0.9238,0.9524)

	Development Time Period	Validation Time Period
WA	0.8840 (0.8611,0.9037)	0.9040 (0.8845,0.9234)
WI	0.8469 (0.8118,0.8820)	0.8703 (0.8374,0.9032)
WY	0.8261 (0.7166,0.8351)	0.8305 (0.7348,0.9020)

Figure 1. Informed Coverage (95% CI) by State ordered by State Informed Coverage for Development Time Period

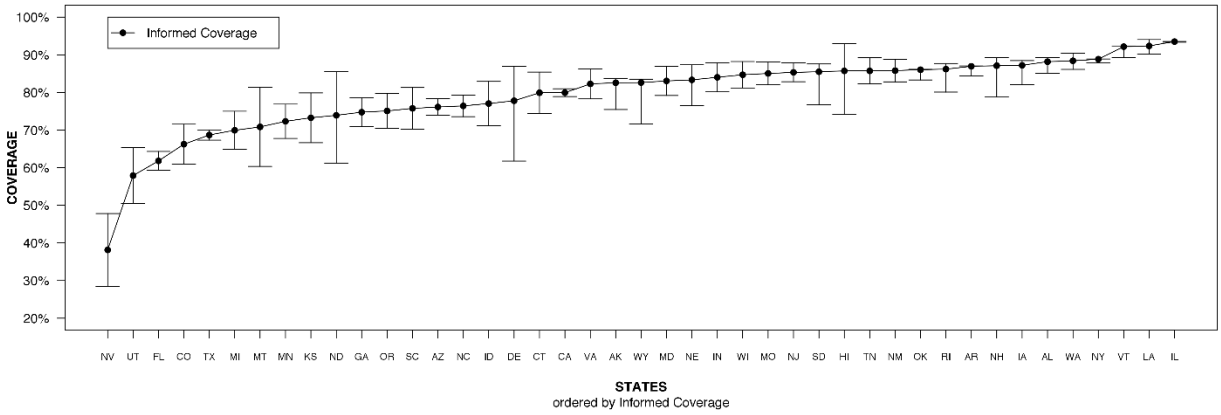
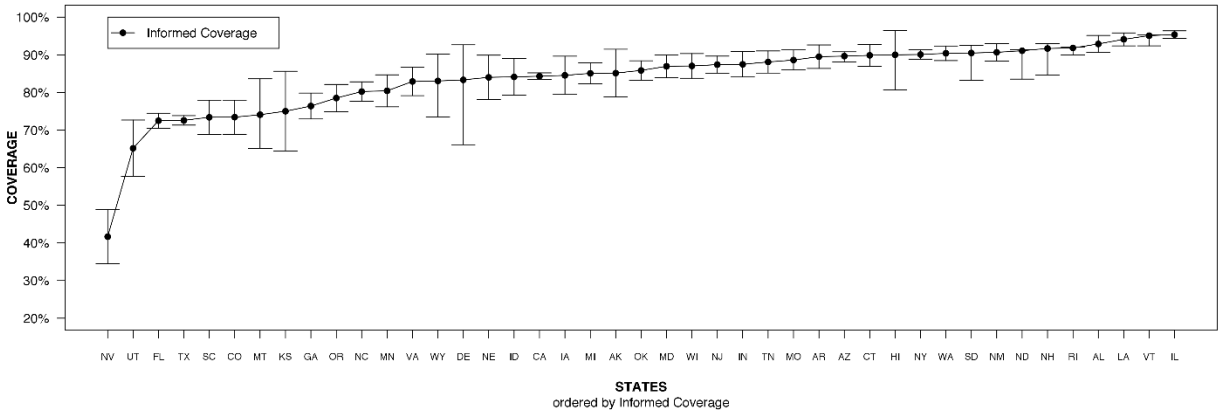


Figure 2. Informed Coverage (95% CI) by State ordered by State Informed Coverage for Validation Time Period



2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)
States were identified with non-overlapping Informed Coverage values, indicating that measure values are variable across the entities measured.

_____

2b2. VALIDITY TESTING

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

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

☒ **Performance measure score**

- ☒ **Empirical validity testing**
- ☐ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*
To examine construct validity, we report Pearson correlations and median absolute errors between the external gold standard of the ACS-based metric and Informed Coverage. In addition to Informed Coverage, there are two additional measures that utilize MAX data utilize to assess the construct validity of Informed Coverage: 1) Duration of first observed enrollment and 2) Continuity Ratio. Duration defines a period of enrollment that begins after new entry to Medicaid/CHIP, birth, or an insurance gap of at least one month, and reports percentage still enrolled at 6, 12, or 18 months. Currently, Duration is utilized voluntarily by states to report to CMS as part of the CHIP Annual Report Template System. Duration is subject to heavy left hand censoring, as continuously enrolled children, the majority of the state, are not eligible for measurement. The Continuity Ratio was developed by Leighton Ku, et al., it calculates the average number of children enrolled per month divided by the number of children enrolled at any point in the year. While the Continuity Ratio has some conceptual resemblance to measures or participation, using the Continuity Ratio as a participation estimate requires the assumption that children who were eligible at any point in the year were eligible the entire year because it makes no adjustments to the denominator for children who enroll for the first time mid-year or who age out of eligibility. Correlations between each of these four measures are presented for the two time periods. It was hypothesized that Informed Coverage would have a higher correlation with the ACS than the other two measures.

Additionally, the Bland-Altman plots were examined for ACS versus Informed Coverage and ACS versus the Continuity Ratio, using each time period.

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

Table 2: Pearson correlations (with 95% CIs) between metrics across states (N=43) in two 18-month time periods, January 2008-June 2009, used with the 2008 ACS and a split sample of the 2009 ACS for metric development, and July 2009-December 2010, used with the remainder of the 2009 ACS split sample and the entire 2010 survey, for validation.

Table 2. Pearson Correlation Results

		Continuity Ratio	Duration	ACS
Informed Coverage	**Development**	0.77[d] (0.61,0.87)	0.52[c] (0.26,0.71)	0.86[d] (0.76,0.92)
	Validation	0.73[d] (0.55,0.85)	0.49[c] (0.23,0.69)	0.86[d] (0.76,0.92)
Continuity Ratio	**Development**	1	0.83[d] (0.70,0.90)	0.75[d] (0.57,0.85)
	Validation		0.71[d] (0.53,0.83)	0.67[d] (0.46,0.80)
Duration	**Development**		1	0.46[b] (0.19,0.67)
	Validation			0.51[c] (0.24,0.70)

[a] p<0.05, [b] p<0.01, [c] p<0.001, [d] p<0.0001

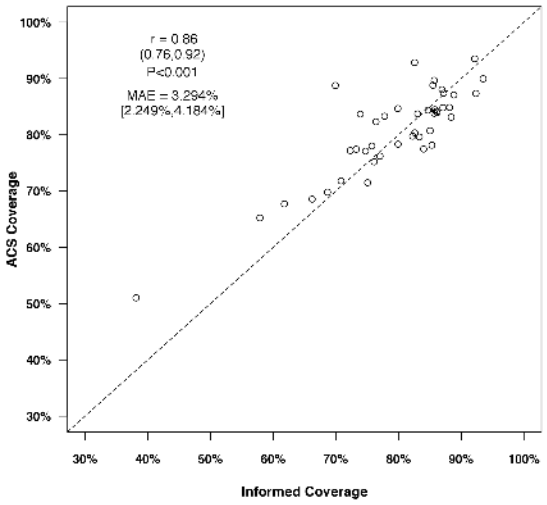
Table 3: Median Absolute Errors (with 95% CIs) between the ACS and Informed Coverage/Continuity Ratio across states (N=43) in two 18-month time periods, January 2008-June 2009, used with the 2008 ACS and a split sample of the 2009 ACS for metric development, and July 2009-December 2010, used with the remainder of the 2009 ACS split sample and the entire 2010 survey, for validation. Note: Duration is not included here since it is measuring a different construct. However, state performance on duration can be compared to state ACS performance through correlation (as are the other metrics).

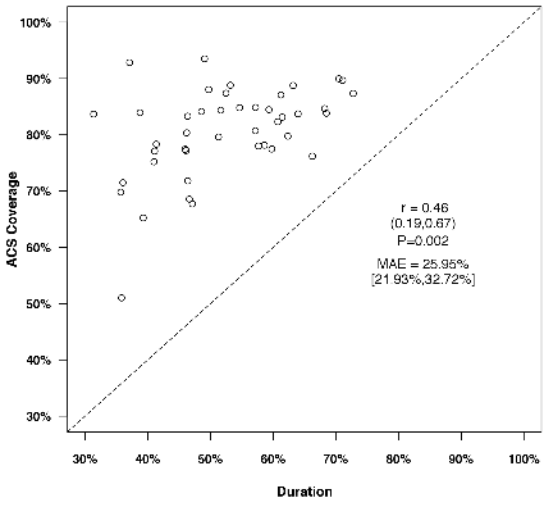
Table 3. Median Absolute Error Results

		Informed Coverage	Continuity Ratio
ACS	**Development**	3.29% (2.24%,4.18%)	6.47% (5.05%,9.34%)
	Validation	3.67%	8.23%

		(2.54%,4.46%)	(7.42%,10.29%)

Figure 3: Correlations and Median Absolute Errors (MAE) examining ACS relative to Informed Coverage, Duration, and Continuity Ratio over 43 States for the development time period. For each measure, we provide correlation plots and median absolute errors. Deleting the outlier state with the lowest ACS coverage, Nevada, did not appreciably change these results.





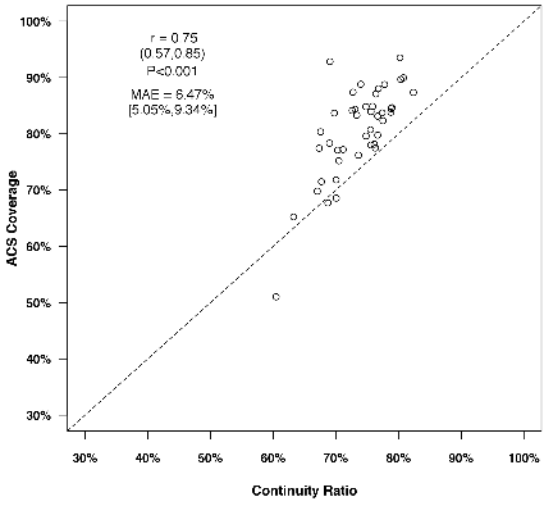
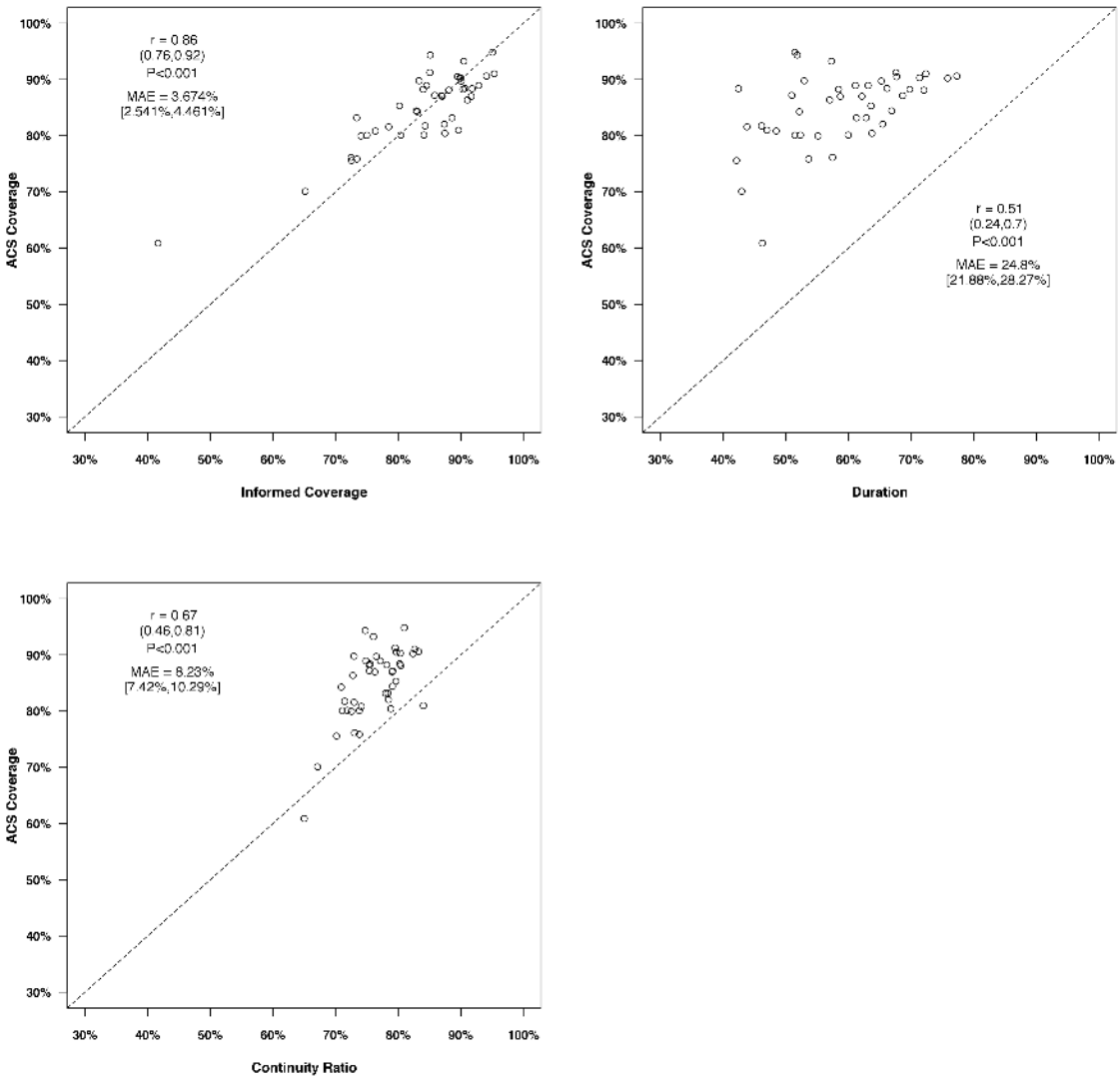


Figure 4: Correlations and Median Absolute Errors (MAE) examining ACS relative to Informed Coverage, Duration, and Continuity Ratio over 43 States for the validation time period. For each measure, we provide correlation plots and median absolute errors. Deleting the outlier state with the lowest ACS coverage, Nevada, did not appreciably change these results.



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

We hypothesized that Informed Coverage would have a higher correlation with ACS than the other measures, which was supported by the results for both the development and validation data. Informed Coverage was well correlated with the ACS metric across the 43 states (r = 0.86 (0.76, 0.92)), and showed similar correlation with ACS in the validation set (r = 0.86 (0.76, 0.92)). The Continuity Ratio and Duration displayed far lower correlations in the two time periods.

The median absolute error (MAE) between ACS and Informed Coverage was calculated. In the development data, the Informed Coverage showed a low MAE of 3.29% with respect to the 2008/2009 ACS survey and was 3.67% in the validation study. The MAE between the ACS and the Continuity Ratio was over twice as large than for Informed Coverage for both the development and validation data sets.

In summary, using the ACS survey as a gold standard, participation rates using Informed Coverage were highly correlated with the survey and error rates were low.

2b3. EXCLUSIONS ANALYSIS
NA ☐ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)
Only children between the ages of 2 and 16 years are included in the intermediate calculation of the Appendicitis Coverage rate. To clarify, the Appendicitis Coverage Rate is calculated using children between the ages of 2 and 16 using age as a continuous number. Therefore, only children that are admitted for an appendectomy the day after their 2nd birthday, up till the day before their 16th birthday would be included. A diagnosis of appendicitis in children 2 years or less is incredibly rare and unreliable. Children 16 years or older were excluded because of concern for censoring due to aging out during 18 month observation window. We did not believe that including 16 to 18 year olds could possibly change the coverage rates to any appreciable amount. We would remind the reader that using the ACS as a gold standard, our median absolute error was less than half that of the alternative measure of continuity created by Ku, and our informed coverage was more highly correlated with ACS than Ku.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)
There were 32,653 inpatient claims for appendicitis or appendectomy in children 0-18 years old between January 1, 2008 and June 30, 2009, of these 5,079 (15.6%) were excluded from the Appendicitis Coverage rate calculation because the patient was either between 2 or less or 16 years or older. The Appendicitis Coverage rate was compared for the patients admitted at an age between 2-16 with those admitted at an age between 0-18 years. The median absolute difference was 0.61% when comparing these two rates. The difference ranged between -2.09%-2.55% across the 43 states used in the analysis.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)
There is minimal difference in the intermediate calculation of the Appendicitis Coverage rate when comparing the value calculated when using patients between the ages of 2-16 and patients between the ages of 0-18 years old.

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

2b4.1. What method of controlling for differences in case mix is used?
☒ **No risk adjustment or stratification**
☐ **Statistical risk model with _risk factors**
☐ **Stratification by _risk categories**
☐ **Other,**

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.
N/A

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.
States need to know whether their participation rates are above or below the participation rates of other states and act accordingly. To perform any risk adjustment would evaluate states using different standards, granting leeway for some states. As all states should strive to improve their process to ensure enrollment of eligible children, risk adjusting would not provide an accurate picture of the state's need to improve its coverage status. States need to adjust their approach to enrolling children based on unadjusted measurements of the problem. Therefore, it is not our intent to risk adjust these rates.

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

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

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

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

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

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

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

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

2b4.9. Results of Risk Stratification Analysis:
N/A

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

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

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*
See section 2a2.2. for a description of how practically meaningful differences are assessed across states.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)
Please see table 1 in this document for testing results.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)
There are practically meaningful differences in the Informed Coverage values across the states tested.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

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

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

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

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

_____

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
Mathematic Policy Research reports for 2008, 2009, and 2010 MAX data report the following anomalies for data needed for the Informed Coverage measure (Overarching Note: DC, HI, MO, ND, PA, UT, and WI were not included in MAX 2008 because the corresponding MSIS files were unavailable or contained significant data problems.)

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

- In a given year, the percent of records with no reported Medicaid enrollment (excluding S-CHIP only enrollees) is evaluated. Values greater than 2.0 percent are above the expected level and are considered anomalous. The following states were found to have data anomalies for 2008 enrollment data: Arizona (21.6%), California (2.8%), Florida (2.1%), Hawaii (3.4%), Michigan (3.7%), Oklahoma (3.6%), Utah (3.6%), and Washington (2.1%). The following states were found to have data anomalies for 2009 enrollment data: California (2.6%), Hawaii (3.5%), Utah (3.9%), and Washington (2.3%). The following states were found to have data anomalies for 2010 enrollment data: Arizona (21.6%), California (2.8%), Florida (2.1%), Hawaii (3.4%), Michigan (3.7%), Oklahoma (3.6%), Utah (3.6%), and Washington (2.1%).
- In a given year, the percent of records with date of birth included in the personal summary file is evaluated. Values greater than 2.0 percent are above the expected level and are considered anomalous. No states were found to have greater than 2.0% of enrollees missing date of birth in 2008, 2009, or 2010.
- In a given year, the percent of records with a primary diagnosis included in the inpatient claims file is evaluated. None of the 43 states used in the analysis were found to have less than 100.0% of annual claims including a primary diagnosis. The following state was found to have less than 100.0% of 2010 claims including a primary diagnosis: Kansas (0.0%).
- In a given year, the percent of records with any procedure code included in the inpatient file is evaluated. Values less than 50.0 percent are below the expected level and are considered anomalous. The following states were found to have less than 50.0% of 2008 claims including a procedure code: Arkansas (48.2%), Georgia (45.1%), and Rhode Island (0.2%). The following states were found to have less than 50.0% of 2009 claims including a procedure code: Arkansas (47.3%), Minnesota (50.0%), Missouri (43.2%), and Rhode Island (0.2%). The following states were found to have less than 50.0% of 2010 claims including a procedure code: Arkansas (49.0%), Connecticut (47.4%), Idaho (45.7%), Kansas (0.0%), Maryland (46.7%), Minnesota (49.8%), Missouri (42.6%), and Rhode Island (0.2%).

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing*

data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data)

For persons to be evaluated in the Informed Coverage measure, there must be evidence of enrollment within the 18-month observation period or the 18-month lookback period when conducting the presumed eligible calculation. Therefore, because there is no evidence of enrollment then these people would not be included at any level of the Informed Coverage measure.

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

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

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

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

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

3c. Data Collection Strategy

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

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.
Informed Coverage is designed to be used with the Medicaid Analytic eXtract (MAX) or similar administrative datasets. However, states and programs do not have consistent reporting standards when contributing to MAX. Some states do not report enrollment data, and none report claims for their state funded (S-CHIP) programs. For children enrolled in Medicaid or Medicaid Expansion (M-CHIP), states report the number of days that a child is enrolled which is used in a decision rule determining whether a child is considered covered. Since this information is not included for any of the states that do report S-CHIP status, children are considered to be "enrolled" for the whole month if they have evidence of S-CHIP enrollment via a monthly indicator in the MAX data. In states that do not report S-CHIP enrollment to MAX, we must assess only the Medicaid and M-CHIP children to estimate enrollment. Additionally, while some states usually provide managed care claims, others do not (Byrd and Verdier 2011; Levinson 2009). Since Informed Coverage requires use of claims data, which may be missing or incomplete in states with high managed care populations, we developed a filter to assess data quality to determine whether Informed Coverage may be implemented in a given state and year. We also analyzed the metric's robustness to unobserved data, in order to be used in states that do not report S-CHIP enrollment data.

Although Medicaid and Medicaid-Expansion (M-CHIP) data is generally complete, some states do not report enrollment data, and no states report claims for their state funded (S-CHIP) programs. For children enrolled in Medicaid or M-CHIP, enrollment is reported as days per month (>14 days of enrollment is considered covered for that month). For states reporting S-CHIP, enrollment status is recorded monthly, via an S-CHIP indicator in the MAX data. In states that do not report S-CHIP enrollment to MAX, we must assess only the Medicaid and M-CHIP children. In order to utilize the Informed Coverage metric, we must trust that states provide bills for children in managed care who develop appendicitis, but managed care claims are sometimes absent from MAX (Byrd and Verdier, 2011; Levinson, 2009). In order to address this problem, we review the managed care data reporting relative to fee-for-service (FFS) or primary care case management (PCCM) systems for inpatient appendicitis claims by state. In each state, over a calendar year, the managed care enrollment rate among children who had appendicitis was compared to that of children without appendicitis to assess whether claims for the managed care children were adequately reported. Each child with appendicitis was matched to 10 children without appendicitis via Mahalanobis distance optimal matching (Rosenbaum, 2010) with a distance matrix that included age and exact matched on gender, the two most clinically relevant risk factors for appendicitis (Addiss, 1990). This generated a control pool of children that had the same gender and very similar, if not identical, age to their matched counterpart. For each child with appendicitis, to avoid bias of retroactive coverage, a point-in-time four months before the date of appendicitis admission was used to determine whether the child was covered via FFS/PCCM or managed care, and the same month was used for their non-appendicitis matched counterpart. In the context of non-inferiority testing (Wellek, 2010), a state was deemed to have insufficient managed care appendicitis claims if the 95% confidence interval for the managed care rate in the appendicitis children minus the managed care rate in the matched controls was completely below -2%. Six states failed to meet this criterion and thus were eliminated from analyses for this time period and dataset: Kentucky, Massachusetts, Mississippi, Ohio, Pennsylvania, and West Virginia. Additionally, Maine and the District of Columbia were found to have excessive quality issues in their inpatient records and were likewise eliminated. The results of this validation process are included in Appendix Table 10.

States with separate Medicaid and CHIP administrations should develop ways to routinely merge their data to enhance the feasibility of the measure and facilitate implementation. In addition, routine inclusion of several specific elements (e.g., reason for enrollment, reason for disenrollment, English proficiency, etc.) would provide useful information. Currently, there is a CMS initiative titled "Transforming the Medicaid Statistical Information System (T-MSIS)," which is designed to assess the feasibility of modifying the existing MSIS system to routinely collect additional elements. Additionally, although managed care data collection is improving, a unified standard for collecting and reporting claims from these programs would greatly enhance the use of MAX data for research and assessment purposes.

Six states were excluded from the current analysis after failing the following criteria: In order to utilize the Informed Coverage metric, we must trust that states provide bills for children in managed care who develop appendicitis, but managed care claims are sometimes absent from MAX (Byrd and Verdier, 2011; Levinson, 2009). In order to address this problem, we reviewed the managed care data reporting relative to fee-for-service (FFS) or primary care case management (PCCM) systems for inpatient appendicitis claims by state. In each state, over a calendar year, the managed care enrollment rate among children who had appendicitis was compared to that of children without appendicitis to assess whether claims for the managed care children were adequately reported. It is important to note that these states were excluded based on the MAX data for 2008. States may have adequate data internally that would allow them to utilize the Informed Coverage measure and may also have adequate MAX data during other time periods that would be sufficient for that time period's analysis. The exclusion of these six states is specific to MAX data for the time period presented.

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

4. Usability and Use

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

4a. Accountability and Transparency

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

4.1. Current and Planned Use

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Reporting Quality Improvement (external benchmarking to organizations) Quality Improvement (Internal to the specific organization) Not in use	

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

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

N/A

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

The measure is new and has not had widespread or targeted distribution to audiences of interest yet.

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

Within the calendar year of any endorsement, information on this metric will be provided to the National Association of Medicaid Directors (NAMD). The aim will be for the Measure Developer to give a webinar to the NAMD, as past webinars have included similar topics such as data analytics to support Medicaid reform. In addition, overviews will be provided to each state's Medicaid department via electronic mail.

Improvement

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

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

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

Our aim is that by using Informed Coverage, states can better estimate participation in their Medicaid and CHIP populations and in turn improve access and quality of care for their children.

4c. Unintended Consequences

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

4c.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**
No unintended consequences have been identified.

4c.2. **Please explain any unexpected benefits from implementation of this measure.**
This measure has not yet been implemented.

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
At this time, performance results, data, and assistance with interpretation have not been provided to those being measured or other users during implementation. Medicaid/CHIP Programs of New Jersey, Pennsylvania, and Massachusetts were all integral to concept development of the Informed Coverage metric. These entities were part of a stakeholder group that provided feedback to the Children's Hospital of Philadelphia as part of their work as the AHRQ PQMP CoE. This included reviewing the metric design and specifications.

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
Over the course of a year, approximately 10 conference calls were held with Medicaid/CHIP programs to present the work on PQMP measures, including work related to Continuity of Insurance, and to incorporate stakeholder input into measure development. Their input was specifically sought for their insight into feasibility and usability. In addition, the following questions drove discussions: 1) How would a coverage metric complement the duration metric in measuring state or plan performance?; 2) What aspects or types of care would suffer with good coverage but poor duration and vice versa?; 3) Given the resources to reform the Medicaid/CHIP system, what changes would you implement to improve coverage? Duration?.

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.
Feedback was obtained via conference calls/presentations with Medicaid/CHIP programs.

4d2.2. **Summarize the feedback obtained from those being measured.**
The NJ Stakeholder Representative (May 14, 2012 discussion) believed that a system which is insensitive to gaps should not be used at all. They also believed that only the proposed metric should be used. The MA Stakeholder Representative (April 23, 2012 discussion) was in favor of the use of an 18-month observation window and recognized that a state which focused on newly enrolling and re-enrolling disenrollees would not be rewarded or have its success tracked under the current CMS metric.

4d2.3. **Summarize the feedback obtained from other users**
N/A

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
Medicaid/CHIP programs provided input into measure development which has been included in development and revisions to create the current version of the measure.

5. Comparison to Related or Competing Measures

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

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

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

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

5a. Harmonization of Related Measures
The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

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

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

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

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Appendix_Silber_3154_v17-636174843291356508.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** The Children´s Hospital of Philadelphia
Co.2 **Point of Contact:** Jeffrey, Silber, silber@email.chop.edu, 215-590-2540-
Co.3 **Measure Developer if different from Measure Steward:** The Children´s Hospital of Philadelphia
Co.4 **Point of Contact:** Jeffrey, Silber, silber@email.chop.edu, 215-590-2540-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

The following people were part of the Children's Hospital of Philadelphia PQMP Center of Excellence and contributed in conceptualizing the measure: Scott A. Lorch, MD, MSCE, Rose E. Calixte, PhD, Ashley E. Zeigler, BA, Jeanhee Moon, PhD, Christopher B. Forrest, MD, PhD, Susmita Pati, MD, MPH, and Shawna R. Calhoun, MPH. In addition, Russell Localio, PhD, Wei Wang, PhD, Justin Ludwig, MA, and Joseph G. Reiter, MS contributed to the conceptualization of the measure, as well as the statistical design.

Scott A. Lorch, MD, MSCE is associated with Center for Outcomes Research, The Children's Hospital of Philadelphia, Philadelphia, PA; The Departments of Pediatrics, The University of Pennsylvania School of Medicine, Philadelphia, PA; The Leonard Davis Institute of Health Economics, The University of Pennsylvania, Philadelphia, PA; and Center for Perinatal and Pediatric Health Disparities Research, The Children's Hospital of Philadelphia, Philadelphia, PA.

Rose E. Calixte, PhD and Susmita Pati, MD, MPH are associated with Division of Primary Care Pediatrics, Stony Brook University School of Medicine & Stony Brook Children's Hospital, Stony Brook, NY.

Ashley E. Zeigler, BA, Shawna R. Calhoun, MPH, Wei Wang, PhD, Justin Ludwig, MA, and Joseph G. Reiter, MS are associated with Center for Outcomes Research, The Children's Hospital of Philadelphia, Philadelphia, PA.

Jeanhee Moon, PhD is associated with Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Christopher B. Forrest, MD, PhD is associated with The Departments of Pediatrics, The University of Pennsylvania School of Medicine, Philadelphia, PA; The Leonard Davis Institute of Health Economics, The University of Pennsylvania, Philadelphia, PA; and Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Russell Localio, PhD is associated with Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA; and Department of Biostatistics and Epidemiology, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2013

Ad.3 **Month and Year of most recent revision:** 08, 2014

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

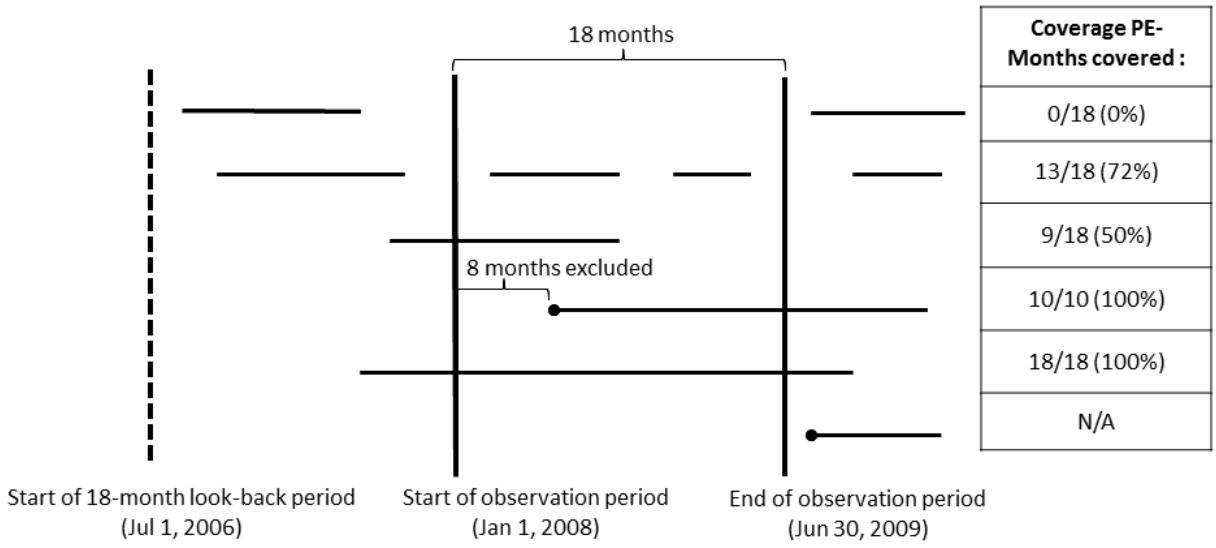
Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:**

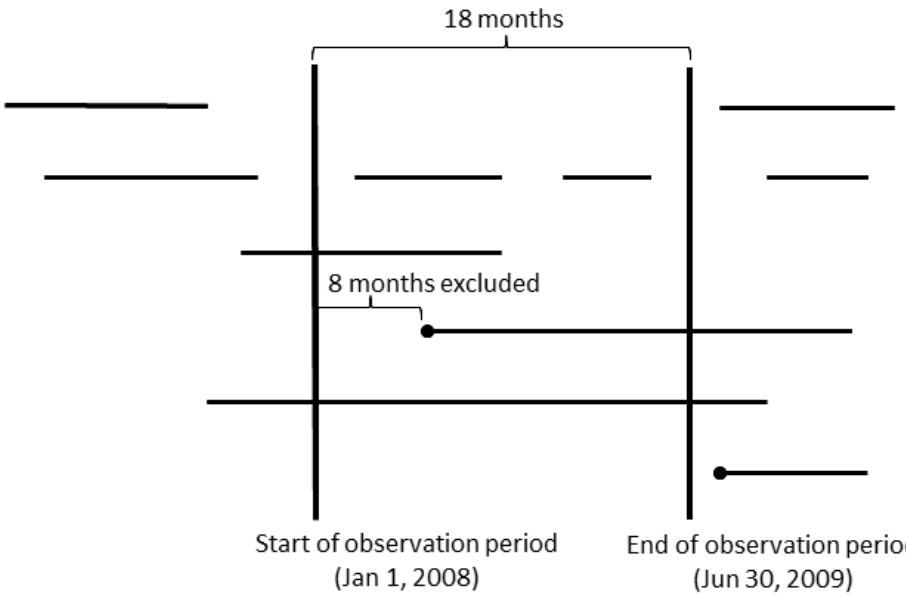
Ad.8 **Additional Information/Comments:**

Figure 1: Coverage PE (Presumed Eligible)



Coverage PE- Months covered :
0/18 (0%)
13/18 (72%)
9/18 (50%)
10/10 (100%)
18/18 (100%)
N/A

Figure 2: Coverage PI (Presumed Ineligible)



Coverage PI- Months covered :
N/A
13/15 (87%)
9/18 (50%)
10/10 (100%)
18/18 (100%)
N/A

Figure 3: Derivation of PE'

Derivation of PE'

Let a = number of appendectomy patients in the state which lacked coverage during the entire lookback until the 4th month before the appendectomy.

Let b = the total number of appendectomy patients

Then the proportion of appendectomy patients without coverage before the appendectomy is the ratio

$$K = \frac{a}{b}$$

This ratio represents the likelihood that any patient would lack coverage, and would not be observed in the claims data. To determine the expected number of patients in the state that would not be observed (X) , we can solve the equation for X:

$$K = \frac{X}{N_{PE} + X}$$

$$X = K * N_{PE} + K * X$$

$$X * (1 - K) = K * N_{PE}$$

$$X = N_{PE}\left(\frac{K}{1 - K}\right)$$

PE is the sum of the covered months divided by the sum of the eligible months. The sum of the months covered is the same as the average number of months covered per person (C) times the number of people (N_{PE}). Similarly, the sum of the months eligible is the same as the average number of months eligible (E) times the number of people (N_{PE}). So PE is the ratio of the average number of covered months divided by the average number of eligible months:

$$PE = \frac{\sum Coverd\ Months}{\sum Eligible\ Months} = \frac{C * N_{PE}}{E * N_{PE}} = \frac{C}{E}$$

The adjusted presumed eligible coverage (PE') adds in eligibility for each presumed unobserved person, giving each unobserved person the average number of eligible months observed in the data:

$$PE' = \frac{C * N_{PE}}{E * N_{PE} + E * X}$$

Which reduces to:

$$PE' = \frac{C * N_{PE}}{E * N_{PE} + E * N_{PE}\left(\frac{K}{1-K}\right)}$$

$$= \frac{C}{E + E\left(\frac{K}{1-K}\right)}$$

Consider the ratio of PE' / PE:

$$\frac{PE'}{PE} = \frac{\frac{C}{E + E\left(\frac{K}{1-K}\right)}}{\frac{C}{E}}$$

$$= \frac{C}{E + E\left(\frac{K}{1-K}\right)} * \frac{E}{C}$$

$$= \frac{1}{1 + \left(\frac{K}{1-K}\right)}$$

$$= \frac{1}{\frac{1-K+K}{1-K}}$$

$$= 1 - K$$

So PE' can be written as:

$$PE' = PE * (1 - K)$$

Variance Formulas for Informed Coverage:

$$Var(ACR) = \frac{ACR * (1 - ACR)}{N_{App}}$$

N_{App} is the number of appendectomy patients in the state

ACR is the state's Appendectomy Coverage Ratio (the proportion of appendectomy patients having insurance coverage prior to developing appendicitis)

$$Var(PE) = \frac{PE * (1 - PE)}{N_{PE}}$$

$$Var(PI) = \frac{PI * (1 - PI)}{N_{PI}}$$

$$Var(K) = \frac{K * (1 - K)}{N_{App}}$$

K is the state's proportion of appendectomy patients without any previous evidence of coverage

$$informed\ PE = \ PE' = PE * (1 - \ K)$$

$$Var(PE') = Var(PE) + Var(PE) * Var(K) + Var(PE) * K^2 + Var(K) * PE^2$$

$$= \frac{PE * (1 - PE)}{N_{PE}} + \frac{PE * (1 - PE)}{N_{PE}} * \frac{K * (1 - K)}{N_{App}} + \frac{PE * (1 - PE)}{N_{PE}} * K^2 + \frac{K * (1 - K)}{N_{App}} * PE^2$$

Figure 4: Informed Coverage Formula

The formula for Informed Coverage when PE' < ACR < PI:

$$Informed\ Coverage = PE' * w + PI * (1 - w)$$

Where,
PE' is the state's Coverage PE'
PI is the state's Coverage PI
ACR is the state's Appendectomy Coverage Ratio (the proportion of appendectomy patients having insurance coverage prior to developing appendicitis)
w = (PI – ACR) / (PI – PE'),

When ACR ≤ PE', Informed Coverage = PE'
When ACR ≥ PI, Informed Coverage = PI

The 95% CI for ACR, PE', PI, K, and IPE are determined using the normal approximation formula:

$$estimate \pm 1.96 * standard\_error(estimate)$$

The endpoints (lower, upper) of the 95% Confidence Interval for Informed Coverage are given by:

$$lower\ 95\%\ CI\ endpoint = \min(\max(lower\ PE', lower\ ACR), lower\ PI)$$

$$upper\ 95\%\ CI\ endpoint = \max(\min(upper\ PI, upper\ ACR), upper\ PE')$$

For stratified analyses, the weight used in the Informed Coverage calculation is specific to the state and is the same for all strata within that state. The 95% confidence interval is determined using the same approach as used for the overall state, but now using the strata specific PE' and PI, and centering the interval at the point estimate for the strata's Informed Coverage with the variance of the ACR term equal to the state's overall appendectomy coverage rate's variance.

butions of Intermediate Calculation and final Informed Coverage Rates

Mean	Std Dev	Min	Max	IQR	Deciles							
					10%	20%	30%	40%	50%	60%	70%	80%
0.7319	0.0635	0.5868	0.8366	0.6805, 0.7807	0.6384	0.6729	0.6991	0.7211	0.7451	0.7554	0.7686	0.784
0.6473	0.0962	0.3206	0.8041	0.5826, 0.7072	0.5112	0.5716	0.6021	0.6452	0.6661	0.6858	0.7039	0.712
0.8763	0.0352	0.8008	0.9398	0.8517, 0.9030	0.8234	0.8460	0.8541	0.8669	0.8830	0.8884	0.9003	0.906
0.7961	0.1049	0.3814	0.9492	0.7474, 0.8580	0.6623	0.7326	0.7576	0.7778	0.8261	0.8469	0.8571	0.861
0.7949	0.1035	0.3814	0.9350	0.7474, 0.8580	0.6623	0.7326	0.7576	0.7778	0.8261	0.8469	0.8571	0.861

Table 2: Number and percentage of children included in the Coverage PE and Coverage PI intermediate calculations (and thus Informed Coverage), out of all children present in the MAX dataset for the years 2008-2009

State	Total N	Coverage PE		Coverage PI	
		N	%	N	%
AK	107847	102455	95.00%	87499	81.13%

AL	605911	579020	95.56%	505744	83.47%
AR	555216	526155	94.77%	475791	85.69%
AZ	1045389	1007211	96.35%	883266	84.49%
CA	6163023	5799593	94.10%	4843863	78.60%
CO	502683	486666	96.81%	425653	84.68%
CT	339837	322906	95.02%	291392	85.74%
DC	102996	97350	94.52%	87164	84.63%
DE	111371	106565	95.68%	94165	84.55%
FL	2218973	2133294	96.14%	1814216	81.76%
GA	1561756	1502114	96.18%	1286398	82.37%
HI	145754	139056	95.40%	123582	84.79%
IA	330476	315471	95.46%	270317	81.80%
ID	202107	194694	96.33%	174039	86.11%
IL	1849889	1770535	95.71%	1623778	87.78%
IN	859129	824018	95.91%	732893	85.31%
KS	295234	284865	96.49%	235591	79.80%
LA	939972	890868	94.78%	774686	82.42%
MD	636309	606278	95.28%	540975	85.02%
MI	1338672	1272596	95.06%	1149476	85.87%
MN	539670	514201	95.28%	438407	81.24%
MO	786171	748346	95.19%	661986	84.20%
MS	523197	499677	95.50%	426188	81.46%
MT	93077	89398	96.05%	77264	83.01%
NC	1303218	1248773	95.82%	1133459	86.97%
ND	56320	54487	96.75%	45911	81.52%
NE	208002	199602	95.96%	172050	82.72%
NH	114015	108554	95.21%	97182	85.24%
NJ	852541	817617	95.90%	710394	83.33%
NM	409034	390749	95.53%	353989	86.54%
NV	241932	235600	97.38%	187935	77.68%
NY	2415869	2291206	94.84%	1957955	81.05%
OK	636032	610104	95.92%	541334	85.11%
OR	415204	399009	96.10%	341723	82.30%
RI	121608	114904	94.49%	102351	84.16%
SC	660279	626265	94.85%	549402	83.21%
SD	109728	105205	95.88%	91644	83.52%
TN	911419	865104	94.92%	780665	85.65%
TX	3838762	3717428	96.84%	3177289	82.77%
UT	280729	273658	97.48%	221775	79.00%
VA	729740	698245	95.68%	627597	86.00%
VT	80895	76070	94.04%	70204	86.78%
WA	851837	805690	94.58%	720429	84.57%
WI	645218	612847	94.98%	551394	85.46%
WY	71174	68401	96.10%	56151	78.89%

Table 3. Intermediate Calculations (95% CI) with Informed Coverage (95% CI) ; Weights are also provided

	Coverage PE	Informed PE	Coverage PI	Appendicitis Coverage	Coverage PE Weight	Coverage PI Weight	Informed Coverage
AK	0.6736 (0.6707,0.8323)	0.6303 (0.5992,0.6615)	0.8348 (0.8323,0.8372)	0.8257 (0.7545,0.8969)	0.0445	0.9555	0.8257 (0.7545,0.8372)
AL	0.7451 (0.7440,0.8904)	0.7055 (0.6894,0.7217)	0.8913 (0.8904,0.8922)	0.8816 (0.8505,0.9128)	0.0520	0.9480	0.8816 (0.8505,0.8922)

AR	0.7686 (0.7675,0.8685)	0.7112 (0.6910,0.7313)	0.8694 (0.8685,0.8704)	0.8763 (0.8435,0.9090)	0.0000	1.0000	0.8694 (0.8435,0.8704)
AZ	0.7028 (0.7020,0.8453)	0.6021 (0.5893,0.6150)	0.8460 (0.8453,0.8468)	0.7612 (0.7391,0.7834)	0.3477	0.6523	0.7612 (0.7391,0.7834)
CA	0.6729 (0.6725,0.8481)	0.6047 (0.5994,0.6100)	0.8484 (0.8481,0.8488)	0.7993 (0.7889,0.8097)	0.2015	0.7985	0.7993 (0.7889,0.8097)
CO	0.7003 (0.6991,0.8526)	0.5534 (0.5213,0.5854)	0.8537 (0.8526,0.8547)	0.6623 (0.6092,0.7154)	0.6373	0.3627	0.6623 (0.6092,0.7154)
CT	0.8094 (0.8081,0.9190)	0.6864 (0.6465,0.7263)	0.9200 (0.9190,0.9210)	0.7990 (0.7440,0.8540)	0.5180	0.4820	0.7990 (0.7440,0.8540)
DE	0.7362 (0.7336,0.8648)	0.6953 (0.6174,0.7733)	0.8669 (0.8648,0.8691)	0.7778 (0.5857,0.9698)	0.5196	0.4804	0.7778 (0.6174,0.8691)
FL	0.6805 (0.6799,0.8557)	0.5076 (0.4923,0.5230)	0.8562 (0.8557,0.8568)	0.6180 (0.5928,0.6431)	0.6836	0.3164	0.6180 (0.5928,0.6431)
GA	0.6724 (0.6717,0.8358)	0.5592 (0.5368,0.5816)	0.8364 (0.8358,0.8371)	0.7474 (0.7088,0.7860)	0.3210	0.6790	0.7474 (0.7088,0.7860)
HI	0.8029 (0.8008,0.9264)	0.7799 (0.7356,0.8243)	0.9278 (0.9264,0.9293)	0.8571 (0.7412,0.9731)	0.4779	0.5221	0.8571 (0.7412,0.9293)
IA	0.7186 (0.7171,0.8821)	0.6485 (0.6159,0.6812)	0.8833 (0.8821,0.8846)	0.8720 (0.8208,0.9231)	0.0485	0.9515	0.8720 (0.8208,0.8846)
ID	0.7606 (0.7588,0.8838)	0.6365 (0.5970,0.6759)	0.8852 (0.8838,0.8867)	0.7704 (0.7115,0.8293)	0.4616	0.5384	0.7704 (0.7115,0.8293)
IL	0.8366 (0.8360,0.9347)	0.8041 (0.7958,0.8124)	0.9350 (0.9347,0.9354)	0.9449 (0.9332,0.9565)	0.0000	1.0000	0.9350 (0.9332,0.9354)
IN	0.7699 (0.7690,0.8996)	0.7072 (0.6850,0.7295)	0.9003 (0.8996,0.9010)	0.8401 (0.8014,0.8788)	0.3117	0.6883	0.8401 (0.8014,0.8788)
KS	0.6384 (0.6366,0.8218)	0.5716 (0.5423,0.6008)	0.8234 (0.8218,0.8249)	0.7326 (0.6664,0.7987)	0.3607	0.6393	0.7326 (0.6664,0.7987)
LA	0.7975 (0.7967,0.9393)	0.7555 (0.7410,0.7699)	0.9398 (0.9393,0.9403)	0.9235 (0.9020,0.9450)	0.0885	0.9115	0.9235 (0.9020,0.9403)
MD	0.7828 (0.7818,0.9071)	0.6850 (0.6582,0.7117)	0.9078 (0.9071,0.9086)	0.8306 (0.7918,0.8693)	0.3468	0.6532	0.8306 (0.7918,0.8693)
MI	0.7945 (0.7938,0.9060)	0.6512 (0.6175,0.6849)	0.9066 (0.9060,0.9071)	0.6994 (0.6488,0.7499)	0.8114	0.1886	0.6994 (0.6488,0.7499)
MN	0.6988 (0.6976,0.8615)	0.5767 (0.5497,0.6037)	0.8625 (0.8615,0.8635)	0.7231 (0.6776,0.7686)	0.4878	0.5122	0.7231 (0.6776,0.7686)
MO	0.7554 (0.7544,0.8823)	0.7039 (0.6876,0.7201)	0.8830 (0.8823,0.8838)	0.8504 (0.8200,0.8808)	0.1822	0.8178	0.8504 (0.8200,0.8808)
MT	0.6991 (0.6963,0.8517)	0.5826 (0.5224,0.6429)	0.8541 (0.8517,0.8564)	0.7083 (0.6033,0.8133)	0.5369	0.4631	0.7083 (0.6033,0.8133)
NC	0.7926 (0.7919,0.9051)	0.6661 (0.6466,0.6856)	0.9057 (0.9051,0.9062)	0.7641 (0.7356,0.7926)	0.5909	0.4091	0.7641 (0.7356,0.7926)
ND	0.6030 (0.5991,0.8485)	0.5112 (0.4485,0.5739)	0.8517 (0.8485,0.8549)	0.7391 (0.6122,0.8660)	0.3306	0.6694	0.7391 (0.6122,0.8549)
NE	0.7211 (0.7191,0.8700)	0.6452 (0.6045,0.6858)	0.8716 (0.8700,0.8731)	0.8333 (0.7649,0.9017)	0.1688	0.8312	0.8333 (0.7649,0.8731)
NH	0.7672 (0.7647,0.8880)	0.7300 (0.6890,0.7711)	0.8899 (0.8880,0.8919)	0.8710 (0.7875,0.9544)	0.1186	0.8814	0.8710 (0.7875,0.8919)

	Coverage PE	**Informed PE**	**Coverage PI**	**Appendicitis Coverage**	**Coverage PE Weight**	**Coverage PI Weight**	**Informed Coverage**
NJ	0.7493 (0.7484,0.9001)	0.6712 (0.6550,0.6873)	0.9008 (0.9001,0.9014)	0.8533 (0.8284,0.8782)	0.2068	0.7932	0.8533 (0.8284,0.8782)
NM	0.8105 (0.8093,0.9215)	0.7292 (0.7084,0.7500)	0.9224 (0.9215,0.9232)	0.8580 (0.8282,0.8877)	0.3334	0.6666	0.8580 (0.8282,0.8877)

NV	0.5868 (0.5848,0.7990)	0.3206 (0.2625,0.3788)	0.8008 (0.7990,0.8026)	0.3814 (0.2848,0.4781)	0.8734	0.1266	0.3814 (0.2848,0.4781)
NY	0.7334 (0.7328,0.8879)	0.6858 (0.6779,0.6937)	0.8884 (0.8879,0.8888)	0.8928 (0.8792,0.9063)	0.0000	1.0000	0.8884 (0.8792,0.8888)
OK	0.7329 (0.7318,0.8594)	0.6659 (0.6492,0.6825)	0.8603 (0.8594,0.8612)	0.8604 (0.8331,0.8876)	0.0000	1.0000	0.8603 (0.8331,0.8612)
OR	0.6707 (0.6693,0.8314)	0.5951 (0.5724,0.6178)	0.8326 (0.8314,0.8339)	0.7507 (0.7046,0.7969)	0.3447	0.6553	0.7507 (0.7046,0.7969)
RI	0.7553 (0.7528,0.8727)	0.7185 (0.6896,0.7473)	0.8747 (0.8727,0.8768)	0.8618 (0.8008,0.9228)	0.0828	0.9172	0.8618 (0.8008,0.8768)
SC	0.7517 (0.7506,0.8954)	0.6573 (0.6252,0.6895)	0.8962 (0.8954,0.8970)	0.7576 (0.7023,0.8128)	0.5803	0.4197	0.7576 (0.7023,0.8128)
SD	0.7306 (0.7279,0.8713)	0.6835 (0.6387,0.7282)	0.8734 (0.8713,0.8756)	0.8548 (0.7672,0.9425)	0.0980	0.9020	0.8548 (0.7672,0.8756)
TN	0.7976 (0.7968,0.9059)	0.7183 (0.6946,0.7419)	0.9065 (0.9059,0.9072)	0.8571 (0.8225,0.8918)	0.2623	0.7377	0.8571 (0.8225,0.8918)
TX	0.6539 (0.6534,0.8138)	0.5292 (0.5216,0.5368)	0.8143 (0.8138,0.8147)	0.6868 (0.6731,0.7004)	0.4472	0.5528	0.6868 (0.6731,0.7004)
UT	0.5975 (0.5957,0.8106)	0.4472 (0.4084,0.4861)	0.8121 (0.8106,0.8137)	0.5789 (0.5049,0.6529)	0.6390	0.3610	0.5789 (0.5049,0.6529)
VA	0.7846 (0.7836,0.9069)	0.7028 (0.6779,0.7278)	0.9076 (0.9069,0.9083)	0.8225 (0.7828,0.8623)	0.4153	0.5847	0.8225 (0.7828,0.8623)
VT	0.8284 (0.8257,0.9196)	0.8003 (0.7619,0.8386)	0.9216 (0.9196,0.9235)	0.9492 (0.8931,1.0000)	0.0000	1.0000	0.9216 (0.8931,0.9235)
WA	0.7807 (0.7798,0.9023)	0.7412 (0.7289,0.7535)	0.9030 (0.9023,0.9037)	0.8840 (0.8611,0.9069)	0.1175	0.8825	0.8840 (0.8611,0.9037)
WI	0.7673 (0.7662,0.8836)	0.7123 (0.6930,0.7316)	0.8844 (0.8836,0.8852)	0.8469 (0.8118,0.8820)	0.2179	0.7821	0.8469 (0.8118,0.8820)
WY	0.6406 (0.6370,0.8290)	0.5431 (0.4765,0.6097)	0.8320 (0.8290,0.8351)	0.8261 (0.7166,0.9356)	0.0206	0.9794	0.8261 (0.7166,0.8351)

Table 4a. Informed Coverage rates stratified by race/ethnicity, January 1, 2008 – June 30, 2009

	White (Non-Hispanic)	Black (Non-Hispanic)	Hispanic	Other^
AK	0.8183 (0.7471,0.8327)	0.8232 (0.7520,0.8423)	0.8307 (0.7595,0.8525)	0.8305 (0.7593,0.8421)
AL	0.8704 (0.8393,0.8823)	0.8914 (0.8603,0.9018)	0.8885 (0.8574,0.8995)	0.8680 (0.8369,0.8823)
AR	0.8613 (0.8286,0.8628)	0.8971 (0.8643,0.8989)	0.8301 (0.7973,0.8341)	0.8711 (0.8383,0.8729)
AZ	0.7340 (0.7119,0.7562)	0.7691 (0.7470,0.7913)	0.7653 (0.7431,0.7874)	0.7890 (0.7669,0.8112)
CA	0.8141 (0.8037,0.8245)	0.8579 (0.8475,0.8683)	0.8240 (0.8136,0.8344)	0.6478 (0.6374,0.6582)
CO	0.6437 (0.5906,0.6967)	0.6741 (0.6210,0.7272)	0.6874 (0.6343,0.7405)	0.6419 (0.5889,0.6950)
CT	0.7869 (0.7319,0.8419)	0.8045 (0.7496,0.8595)	0.8086 (0.7536,0.8636)	0.7883 (0.7333,0.8433)
DE	0.7642 (0.6015,0.8617)	0.7948 (0.6367,0.8820)	0.7675 (0.6066,0.8634)	0.7546 (0.5873,0.8696)
FL	0.5934 (0.5683,0.6185)	0.6554 (0.6303,0.6805)	0.6029 (0.5778,0.6281)	0.6204 (0.5952,0.6455)

	White (Non-Hispanic)	Black (Non-Hispanic)	Hispanic	Other^
GA	0.7408 (0.7022,0.7793)	0.7594 (0.7208,0.7979)	0.7330 (0.6944,0.7716)	0.7242 (0.6856,0.7628)
HI	0.8237 (0.7078,0.9100)	0.7833 (0.6674,0.8976)	0.8661 (0.7501,0.9375)	0.8667 (0.7508,0.9355)
IA	0.8794 (0.8283,0.8926)	0.8827 (0.8315,0.8978)	0.8561 (0.8050,0.8713)	0.8628 (0.8117,0.8764)
ID	0.7712 (0.7123,0.8301)	0.7864 (0.7275,0.8453)	0.5913 (0.5324,0.6501)	0.7482 (0.6893,0.8071)
IL	0.9177 (0.9061,0.9185)	0.9430 (0.9313,0.9437)	0.9471 (0.9355,0.9478)	0.9309 (0.9192,0.9323)
IN	0.8351 (0.7964,0.8738)	0.8580 (0.8192,0.8967)	0.8332 (0.7945,0.8720)	0.8378 (0.7991,0.8765)
KS	0.7257 (0.6595,0.7918)	0.7609 (0.6948,0.8271)	0.7356 (0.6694,0.8017)	0.7175 (0.6514,0.7837)
LA	0.9052 (0.8837,0.9234)	0.9397 (0.9182,0.9558)	0.8753 (0.8538,0.8955)	0.9097 (0.8882,0.9304)
MD	0.8218 (0.7831,0.8606)	0.8347 (0.7960,0.8735)	0.8465 (0.8077,0.8852)	0.8056 (0.7668,0.8443)
MI	0.6887 (0.6381,0.7392)	0.7301 (0.6796,0.7807)	0.6812 (0.6307,0.7318)	0.6406 (0.5901,0.6912)
MN	0.7228 (0.6773,0.7683)	0.7512 (0.7057,0.7967)	0.6816 (0.6361,0.7271)	0.7242 (0.6787,0.7697)
MO	0.8410 (0.8106,0.8714)	0.8793 (0.8489,0.9081)	0.7835 (0.7530,0.8139)	0.8092 (0.7788,0.8397)
MT	0.7044 (0.5994,0.8094)	0.7040 (0.5990,0.8090)	0.6853 (0.5803,0.7903)	0.7233 (0.6183,0.8283)
NC	0.7503 (0.7218,0.7788)	0.7749 (0.7464,0.8034)	0.7702 (0.7417,0.7987)	0.7694 (0.7409,0.7979)
ND	0.7235 (0.5966,0.8504)	0.7691 (0.6422,0.8881)	0.6844 (0.5575,0.7927)	0.7784 (0.6515,0.8768)
NE	0.8220 (0.7536,0.8637)	0.8872 (0.8188,0.9228)	0.8331 (0.7647,0.8740)	0.8275 (0.7590,0.8717)
NH	0.8707 (0.7872,0.8916)	0.8739 (0.7904,0.9049)	0.8784 (0.7949,0.9043)	0.8599 (0.7765,0.8999)
NJ	0.8555 (0.8306,0.8804)	0.8592 (0.8344,0.8841)	0.8522 (0.8273,0.8771)	0.8300 (0.8051,0.8549)
NM	0.8254 (0.7957,0.8552)	0.8319 (0.8021,0.8617)	0.8737 (0.8439,0.9034)	0.8459 (0.8161,0.8757)
NV	0.3714 (0.2747,0.4680)	0.4226 (0.3259,0.5193)	0.3686 (0.2719,0.4653)	0.3846 (0.2879,0.4812)
NY	0.8872 (0.8737,0.8881)	0.8880 (0.8745,0.8889)	0.8995 (0.8860,0.9003)	0.8684 (0.8549,0.8696)
OK	0.8525 (0.8253,0.8538)	0.8757 (0.8485,0.8780)	0.8661 (0.8388,0.8684)	0.8668 (0.8396,0.8690)
OR	0.7340 (0.6878,0.7802)	0.7758 (0.7296,0.8220)	0.8578 (0.8116,0.9040)	0.6394 (0.5933,0.6856)
RI	0.8775 (0.8166,0.8939)	0.8773 (0.8163,0.8959)	0.8814 (0.8204,0.8965)	0.8446 (0.7837,0.8618)
SC	0.7375 (0.6823,0.7928)	0.7696 (0.7143,0.8249)	0.7531 (0.6978,0.8084)	0.7940 (0.7388,0.8493)

	White (Non-Hispanic)	Black (Non-Hispanic)	Hispanic	Other^
SD	0.8412 (0.7535,0.8642)	0.8405 (0.7528,0.8729)	0.8222 (0.7346,0.8539)	0.8787 (0.7911,0.8980)
TN	0.8457 (0.8110,0.8803)	0.8719 (0.8373,0.9066)	0.8407 (0.8061,0.8753)	0.9200 (0.8854,0.9546)
TX	0.6424 (0.6287,0.6560)	0.6954 (0.6817,0.7090)	0.7031 (0.6895,0.7167)	0.6239 (0.6103,0.6375)
UT	0.5716 (0.4976,0.6456)	0.6314 (0.5574,0.7054)	0.5947 (0.5207,0.6687)	0.5689 (0.4949,0.6429)
VA	0.8123 (0.7725,0.8520)	0.8409 (0.8011,0.8806)	0.8061 (0.7663,0.8458)	0.8076 (0.7679,0.8474)
VT	0.9380 (0.8819,0.9404)	0.9594 (0.9034,0.9713)	0.9498 (0.8937,0.9764)	0.9005 (0.8444,0.9037)
WA	0.8750 (0.8521,0.8964)	0.8919 (0.8690,0.9129)	0.9223 (0.8994,0.9380)	0.8791 (0.8562,0.8993)
WI	0.8385 (0.8035,0.8736)	0.8888 (0.8537,0.9228)	0.8527 (0.8177,0.8878)	0.8213 (0.7862,0.8563)
WY	0.8188 (0.7092,0.8286)	0.8539 (0.7444,0.8786)	0.8332 (0.7237,0.8463)	0.8644 (0.7549,0.8790)

Table 4b. Informed Coverage rates stratified by race/ethnicity, July 1, 2009 – December 31, 2010

	White (Non-Hispanic)	Black (Non-Hispanic)	Hispanic	Other^
AK	0.8358 (0.7724,0.8992)	0.8556 (0.7922,0.9190)	0.8401 (0.7767,0.9035)	0.8527 (0.7893,0.9161)
AL	0.9222 (0.8994,0.9449)	0.9349 (0.9121,0.9577)	0.9305 (0.9077,0.9533)	0.9214 (0.8986,0.9441)
AR	0.8957 (0.8645,0.9269)	0.9105 (0.8793,0.9417)	0.9002 (0.8690,0.9314)	0.8778 (0.8466,0.9090)
AZ	0.8887 (0.8732,0.9020)	0.9089 (0.8933,0.9221)	0.8942 (0.8786,0.9059)	0.9123 (0.8967,0.9235)
CA	0.8454 (0.8365,0.8544)	0.8796 (0.8707,0.8885)	0.8579 (0.8490,0.8669)	0.7640 (0.7550,0.7729)
CO	0.7202 (0.6748,0.7655)	0.7367 (0.6914,0.7820)	0.7435 (0.6982,0.7889)	0.7179 (0.6726,0.7633)
CT	0.8921 (0.8631,0.9212)	0.9001 (0.8711,0.9292)	0.9007 (0.8717,0.9298)	0.8960 (0.8670,0.9251)
DE	0.8198 (0.6477,0.9195)	0.8406 (0.6685,0.9332)	0.8252 (0.6530,0.9244)	0.8079 (0.6357,0.9233)
FL	0.7069 (0.6871,0.7267)	0.7466 (0.7268,0.7664)	0.7143 (0.6944,0.7341)	0.7061 (0.6863,0.7259)
GA	0.7565 (0.7219,0.7911)	0.7660 (0.7315,0.8006)	0.7718 (0.7372,0.8064)	0.7493 (0.7147,0.7838)
HI	0.8817 (0.7887,0.9554)	0.8771 (0.7842,0.9614)	0.8958 (0.8028,0.9626)	0.9060 (0.8131,0.9702)
IA	0.8452 (0.7943,0.8961)	0.8533 (0.8024,0.9042)	0.8328 (0.7819,0.8837)	0.8392 (0.7883,0.8901)
ID	0.8356 (0.7867,0.8846)	0.8393 (0.7903,0.8883)	0.8794 (0.8304,0.9283)	0.8003 (0.7513,0.8493)
IL	0.9493 (0.9392,0.9593)	0.9616 (0.9515,0.9717)	0.9672 (0.9572,0.9773)	0.9549 (0.9448,0.9649)
IN	0.8682 (0.8343,0.9021)	0.8873 (0.8534,0.9212)	0.8643 (0.8304,0.8982)	0.8660 (0.8321,0.8999)
KS	0.7384 (0.6323,0.8445)	0.7591 (0.6530,0.8651)	0.7465 (0.6405,0.8526)	0.7359 (0.6298,0.8420)
LA	0.9284 (0.9111,0.9457)	0.9507 (0.9334,0.9680)	0.9100 (0.8927,0.9273)	0.9276 (0.9103,0.9449)
MD	0.8603 (0.8305,0.8902)	0.8695 (0.8396,0.8993)	0.8754 (0.8456,0.9053)	0.8559 (0.8260,0.8857)
MI	0.8422 (0.8142,0.8702)	0.8660 (0.8380,0.8939)	0.8364 (0.8084,0.8644)	0.8224 (0.7944,0.8504)
MN	0.8012 (0.7585,0.8438)	0.8280 (0.7854,0.8707)	0.7762 (0.7335,0.8188)	0.8078 (0.7651,0.8504)
MO	0.8781 (0.8514,0.9048)	0.9053 (0.8786,0.9320)	0.8424 (0.8157,0.8691)	0.8568 (0.8301,0.8835)
MT	0.7353 (0.6456,0.8307)	0.7387 (0.6457,0.8342)	0.7316 (0.6404,0.8270)	0.7318 (0.6426,0.8272)
NC	0.7936 (0.7676,0.8197)	0.8047 (0.7787,0.8308)	0.8027 (0.7767,0.8287)	0.8000 (0.7740,0.8260)

	White (Non-Hispanic)	Black (Non-Hispanic)	Hispanic	Other^
ND	0.9062 (0.8204,0.9096)	0.9127 (0.8270,0.9260)	0.8675 (0.7817,0.8815)	0.9236 (0.8379,0.9281)
NE	0.8288 (0.7701,0.8875)	0.8721 (0.8134,0.9308)	0.8399 (0.7813,0.8986)	0.8308 (0.7721,0.8895)
NH	0.9164 (0.8464,0.9294)	0.9352 (0.8653,0.9551)	0.9167 (0.8468,0.9345)	0.9252 (0.8552,0.9467)
NJ	0.8724 (0.8499,0.8950)	0.8757 (0.8532,0.8983)	0.8784 (0.8558,0.9009)	0.8609 (0.8383,0.8834)
NM	0.8869 (0.8636,0.9103)	0.8800 (0.8566,0.9033)	0.9150 (0.8917,0.9383)	0.8986 (0.8753,0.9219)
NV	0.4031 (0.3330,0.4903)	0.4258 (0.3517,0.5129)	0.4066 (0.3359,0.4938)	0.3997 (0.3299,0.4869)
NY	0.8976 (0.8853,0.9099)	0.8972 (0.8849,0.9096)	0.9066 (0.8943,0.9189)	0.8899 (0.8776,0.9022)
OK	0.8519 (0.8260,0.8777)	0.8699 (0.8441,0.8958)	0.8648 (0.8390,0.8907)	0.8584 (0.8325,0.8842)
OR	0.7745 (0.7382,0.8107)	0.7931 (0.7569,0.8294)	0.8158 (0.7796,0.8521)	0.7417 (0.7054,0.7780)
RI	0.9225 (0.8806,0.9260)	0.9210 (0.8790,0.9276)	0.9331 (0.8912,0.9369)	0.9150 (0.8731,0.9174)
SC	0.7214 (0.6765,0.7829)	0.7392 (0.6932,0.8008)	0.7147 (0.6701,0.7762)	0.7417 (0.6955,0.8033)
SD	0.8988 (0.8263,0.9218)	0.8912 (0.8187,0.9205)	0.8909 (0.8184,0.9209)	0.9142 (0.8418,0.9345)
TN	0.8743 (0.8450,0.9036)	0.8924 (0.8631,0.9217)	0.8750 (0.8457,0.9043)	0.8679 (0.8386,0.8972)
TX	0.6901 (0.6774,0.7028)	0.7286 (0.7160,0.7413)	0.7340 (0.7213,0.7466)	0.6957 (0.6830,0.7083)
UT	0.6379 (0.5629,0.7129)	0.6818 (0.6068,0.7568)	0.6630 (0.5880,0.7380)	0.6397 (0.5647,0.7147)
VA	0.8204 (0.7829,0.8580)	0.8394 (0.8018,0.8770)	0.8205 (0.7830,0.8581)	0.8082 (0.7707,0.8458)
VT	0.9584 (0.9324,0.9605)	0.9722 (0.9479,0.9821)	0.9520 (0.9080,0.9788)	0.9425 (0.8985,0.9451)
WA	0.8971 (0.8777,0.9165)	0.9052 (0.8858,0.9247)	0.9254 (0.9060,0.9448)	0.9000 (0.8806,0.9195)
WI	0.8671 (0.8342,0.9000)	0.8873 (0.8545,0.9202)	0.8686 (0.8357,0.9015)	0.8561 (0.8232,0.8890)
WY	0.8219 (0.7261,0.8963)	0.8386 (0.7429,0.9262)	0.8334 (0.7377,0.9089)	0.8459 (0.7501,0.9201)

Table 5a. Informed Coverage rates stratified by special healthcare needs status, January 1, 2008 – June 30, 2009

	Chronic Disease	Healthy
AK	0.8498 (0.7786,0.8640)	0.8206 (0.7494,0.8325)
AL	0.8989 (0.8678,0.9091)	0.8757 (0.8446,0.8867)
AR	0.9020 (0.8692,0.9038)	0.8589 (0.8261,0.8600)

	Chronic Disease	Healthy
AZ	0.8048 (0.7827,0.8270)	0.7529 (0.7308,0.7751)
CA	0.8622 (0.8518,0.8726)	0.7890 (0.7786,0.7994)
CO	0.6995 (0.6464,0.7525)	0.6557 (0.6026,0.7088)
CT	0.8019 (0.7469,0.8569)	0.7989 (0.7439,0.8539)
DE	0.8200 (0.6655,0.9003)	0.7658 (0.6039,0.8609)
FL	0.6556 (0.6305,0.6807)	0.6101 (0.5850,0.6352)
GA	0.7937 (0.7551,0.8323)	0.7371 (0.6985,0.7757)
HI	0.8945 (0.7859,0.9530)	0.8452 (0.7292,0.9220)
IA	0.8995 (0.8483,0.9125)	0.8660 (0.8149,0.8790)
ID	0.7974 (0.7386,0.8563)	0.7646 (0.7058,0.8235)
IL	0.9601 (0.9484,0.9608)	0.9280 (0.9163,0.9284)
IN	0.8688 (0.8301,0.9076)	0.8331 (0.7943,0.8718)
KS	0.7634 (0.6973,0.8296)	0.7257 (0.6596,0.7919)
LA	0.9396 (0.9181,0.9550)	0.9186 (0.8971,0.9361)
MD	0.8583 (0.8196,0.8971)	0.8201 (0.7813,0.8588)
MI	0.7401 (0.6896,0.7907)	0.6933 (0.6428,0.7439)
MN	0.7654 (0.7199,0.8108)	0.7146 (0.6691,0.7600)
MO	0.8779 (0.8475,0.9077)	0.8418 (0.8113,0.8722)
MT	0.7535 (0.6485,0.8585)	0.6990 (0.5940,0.8040)
NC	0.7983 (0.7698,0.8268)	0.7541 (0.7256,0.7826)
ND	0.7981 (0.6712,0.8980)	0.7282 (0.6013,0.8474)
NE	0.8628 (0.7944,0.9001)	0.8262 (0.7577,0.8671)
NH	0.9050 (0.8215,0.9247)	0.8635 (0.7801,0.8854)
NJ	0.8869 (0.8620,0.9117)	0.8450 (0.8202,0.8699)

	Chronic Disease	Healthy
NM	0.8827 (0.8529,0.9125)	0.8511 (0.8213,0.8809)
NV	0.4094 (0.3127,0.5061)	0.3771 (0.2804,0.4737)
NY	0.9060 (0.8924,0.9069)	0.8830 (0.8695,0.8835)
OK	0.8906 (0.8634,0.8925)	0.8518 (0.8246,0.8529)
OR	0.7927 (0.7465,0.8389)	0.7429 (0.6967,0.7891)
RI	0.8870 (0.8260,0.9018)	0.8530 (0.7920,0.8689)
SC	0.7866 (0.7313,0.8419)	0.7474 (0.6922,0.8027)
SD	0.8851 (0.7974,0.9064)	0.8483 (0.7607,0.8697)
TN	0.8867 (0.8521,0.9214)	0.8497 (0.8150,0.8843)
TX	0.7381 (0.7244,0.7517)	0.6748 (0.6611,0.6884)
UT	0.6099 (0.5359,0.6839)	0.5751 (0.5011,0.6491)
VA	0.8575 (0.8177,0.8972)	0.8132 (0.7735,0.8529)
VT	0.9458 (0.8897,0.9492)	0.9130 (0.8569,0.9154)
WA	0.9045 (0.8816,0.9231)	0.8805 (0.8576,0.9006)
WI	0.8918 (0.8567,0.9268)	0.8372 (0.8022,0.8723)
WY	0.8585 (0.7489,0.8719)	0.8200 (0.7105,0.8294)

Table 5b. Informed Coverage rates stratified by special healthcare needs status, July 1, 2009 – December 31, 2010		
	Chronic Disease	**Healthy**
AK	0.8707 (0.8073,0.9341)	0.8480 (0.7846,0.9114)
AL	0.9409 (0.9181,0.9636)	0.9252 (0.9024,0.9480)
AR	0.9174 (0.8862,0.9486)	0.8888 (0.8576,0.9201)
AZ	0.9310 (0.9154,0.9420)	0.8906 (0.8750,0.9025)
CA	0.8905 (0.8816,0.8995)	0.8366 (0.8277,0.8455)
CO	0.7570 (0.7117,0.8024)	0.7311 (0.6858,0.7765)
CT	0.9027 (0.8737,0.9318)	0.8984 (0.8694,0.9275)
DE	0.8529 (0.6807,0.9467)	0.8289 (0.6568,0.9224)
FL	0.7482 (0.7283,0.7680)	0.7210 (0.7012,0.7408)
GA	0.7951 (0.7605,0.8296)	0.7579 (0.7233,0.7925)
HI	0.9153 (0.8223,0.9784)	0.8962 (0.8032,0.9615)
IA	0.8695 (0.8187,0.9204)	0.8415 (0.7906,0.8924)
ID	0.8596 (0.8107,0.9086)	0.8382 (0.7892,0.8871)
IL	0.9711 (0.9611,0.9812)	0.9496 (0.9395,0.9597)
IN	0.8959 (0.8620,0.9298)	0.8704 (0.8366,0.9043)
KS	0.7673 (0.6613,0.8734)	0.7471 (0.6410,0.8532)
LA	0.9502 (0.9328,0.9675)	0.9388 (0.9215,0.9561)
MD	0.8874 (0.8575,0.9172)	0.8642 (0.8344,0.8940)
MI	0.8702 (0.8422,0.8982)	0.8484 (0.8204,0.8763)
MN	0.8403 (0.7976,0.8829)	0.7981 (0.7554,0.8408)
MO	0.9080 (0.8813,0.9347)	0.8805 (0.8538,0.9072)
MT	0.7664 (0.6741,0.8618)	0.7367 (0.6473,0.8322)
NC	0.8248 (0.7988,0.8508)	0.7968 (0.7708,0.8228)

	Chronic Disease	Healthy
ND	0.9415 (0.8557,0.9474)	0.9061 (0.8204,0.9087)
NE	0.8619 (0.8032,0.9206)	0.8359 (0.7772,0.8945)
NH	0.9435 (0.8736,0.9573)	0.9121 (0.8421,0.9248)
NJ	0.8949 (0.8723,0.9174)	0.8699 (0.8473,0.8924)
NM	0.9255 (0.9022,0.9488)	0.9023 (0.8790,0.9256)
NV	0.4358 (0.3598,0.5229)	0.4148 (0.3427,0.5019)
NY	0.9190 (0.9067,0.9313)	0.8965 (0.8842,0.9088)
OK	0.8797 (0.8539,0.9056)	0.8538 (0.8279,0.8796)
OR	0.8131 (0.7768,0.8493)	0.7811 (0.7448,0.8173)
RI	0.9386 (0.8967,0.9417)	0.9124 (0.8705,0.9143)
SC	0.7521 (0.7053,0.8137)	0.7287 (0.6834,0.7903)
SD	0.9256 (0.8531,0.9477)	0.9012 (0.8287,0.9221)
TN	0.8973 (0.8680,0.9266)	0.8775 (0.8482,0.9068)
TX	0.7635 (0.7509,0.7762)	0.7187 (0.7060,0.7313)
UT	0.6768 (0.6018,0.7518)	0.6493 (0.5743,0.7244)
VA	0.8508 (0.8132,0.8883)	0.8244 (0.7869,0.8620)
VT	0.9666 (0.9384,0.9695)	0.9463 (0.9087,0.9481)
WA	0.9203 (0.9008,0.9397)	0.9018 (0.8824,0.9212)
WI	0.9013 (0.8684,0.9341)	0.8652 (0.8323,0.8980)
WY	0.8609 (0.7651,0.9374)	0.8263 (0.7305,0.8977)

Table 6a. Informed Coverage rates, January 1, 2008 – June 30, 2009, stratified by average income determined at the level of a 5-digit zip code: average income.

	Lowest Income	2nd Quartile	3rd Quartile	Highest Income	Missing
AK	-	-	0.7756 (0.2016,1.0000)	-	0.8257 (0.7545,0.8372)
AL	0.4752 (0.4148,0.5064)	0.5094 (0.4514,0.5405)	0.5350 (0.4773,0.5661)	0.5146 (0.4567,0.5457)	0.8822 (0.8510,0.8926)

	Lowest Income	2nd Quartile	3rd Quartile	Highest Income	Missing
AR	0.4870 (0.4045,0.5198)	0.4721 (0.3932,0.5049)	0.4624 (0.3849,0.4952)	0.4027 (0.3231,0.4354)	0.8699 (0.8372,0.8709)
AZ	-	-	-	-	0.7612 (0.7391,0.7834)
CA	0.4881 (0.4777,0.4985)	0.4949 (0.4845,0.5053)	0.4953 (0.4849,0.5057)	0.5041 (0.4937,0.5145)	0.7997 (0.7893,0.8101)
CO	0.3536 (0.3005,0.4067)	0.3358 (0.2827,0.3889)	0.3511 (0.2980,0.4041)	0.3445 (0.2915,0.3976)	0.6634 (0.6104,0.7165)
CT	0.4322 (0.3772,0.4872)	0.4706 (0.4156,0.5256)	0.4355 (0.3806,0.4905)	0.4958 (0.4408,0.5508)	0.8012 (0.7463,0.8562)
DE	-	-	-	-	0.7778 (0.6174,0.8691)
FL	0.3637 (0.3385,0.3888)	0.3685 (0.3434,0.3937)	0.3553 (0.3302,0.3804)	0.3452 (0.3201,0.3703)	0.6196 (0.5945,0.6447)
GA	0.6187 (0.3042,0.8433)	0.5653 (0.2600,0.6039)	0.7727 (0.5696,0.9393)	0.7556 (0.7170,0.7942)	0.7474 (0.7088,0.7860)
HI	-	-	-	-	0.8571 (0.7412,0.9293)
IA	0.5220 (0.4109,0.5731)	0.5853 (0.4809,0.6364)	0.5473 (0.4474,0.5984)	0.5603 (0.4545,0.6114)	0.8723 (0.8211,0.8849)
ID	0.7670 (0.7081,0.8259)	0.7709 (0.7120,0.8298)	0.7765 (0.7176,0.8354)	0.7655 (0.7066,0.8244)	0.7900 (0.7311,0.8489)
IL	0.9344 (0.9227,0.9352)	0.9388 (0.9272,0.9396)	0.9385 (0.9268,0.9392)	0.9291 (0.9175,0.9299)	0.9252 (0.9135,0.9276)
IN	0.6003 (0.5616,0.6391)	0.5617 (0.5230,0.6004)	0.5612 (0.5225,0.6000)	0.5600 (0.5212,0.5987)	0.8423 (0.8035,0.8810)
KS	0.5546 (0.4885,0.6208)	0.5251 (0.4590,0.5913)	0.5270 (0.4608,0.5931)	0.4967 (0.4305,0.5628)	0.7329 (0.6668,0.7991)
LA	0.9408 (0.9193,0.9559)	0.9309 (0.9095,0.9463)	0.9262 (0.9047,0.9433)	0.9161 (0.8946,0.9342)	0.8163 (0.7948,0.8378)
MD	0.4656 (0.2590,0.5059)	0.7723 (0.7335,0.8110)	0.8037 (0.7624,0.8424)	0.7101 (0.6402,0.7489)	0.8306 (0.7918,0.8693)
MI	0.2922 (0.2417,0.3428)	0.3081 (0.2575,0.3587)	0.2807 (0.2301,0.3312)	0.3027 (0.2521,0.3532)	0.7001 (0.6495,0.7506)
MN	0.5884 (0.5429,0.6339)	0.6507 (0.6052,0.6961)	0.6782 (0.6327,0.7236)	0.5749 (0.5294,0.6203)	0.7252 (0.6797,0.7707)
MO	0.7146 (0.6842,0.7450)	0.7696 (0.7391,0.8000)	0.7075 (0.6771,0.7380)	0.7524 (0.7219,0.7828)	0.8509 (0.8205,0.8814)
MT	0.7230 (0.6180,0.8280)	0.7179 (0.6129,0.8229)	0.7090 (0.6040,0.8140)	0.7036 (0.5986,0.8086)	0.5307 (0.4257,0.6357)
NC	0.7832 (0.7547,0.8117)	0.7774 (0.7489,0.8059)	0.7665 (0.7380,0.7950)	0.7502 (0.7217,0.7787)	0.6712 (0.6427,0.6997)
ND	0.7675 (0.6406,0.8779)	0.7341 (0.6072,0.8585)	0.7426 (0.6157,0.8612)	0.7486 (0.6217,0.8698)	0.5702 (0.4433,0.6971)
NE	0.4175 (0.1117,0.4859)	0.5496 (0.3105,0.6180)	0.6791 (0.2456,0.7475)	0.4518 (0.1535,0.5202)	0.8334 (0.7650,0.8732)
NH	0.8827 (0.7992,0.9045)	0.8780 (0.7946,0.9003)	0.8683 (0.7848,0.8910)	0.8551 (0.7717,0.8801)	0.8613 (0.7778,0.8905)
NJ	0.8661 (0.8412,0.8909)	0.8652 (0.8403,0.8900)	0.8450 (0.8201,0.8698)	0.8416 (0.8167,0.8665)	0.7419 (0.7170,0.7668)

	Lowest Income	2nd Quartile	3rd Quartile	Highest Income	Missing
NM	0.7607 (0.7023,0.7905)	0.6536 (0.4333,0.8136)	0.8421 (0.7159,0.9554)	0.8699 (0.8292,0.9430)	0.8580 (0.8282,0.8877)
NV	-	-	0.3019 (0.0000,1.0000)	0.1483 (0.0000,0.6390)	0.3815 (0.2848,0.4781)
NY	0.9019 (0.8884,0.9027)	0.8950 (0.8815,0.8959)	0.8883 (0.8747,0.8891)	0.8744 (0.8608,0.8753)	0.7092 (0.6957,0.7154)
OK	0.4725 (0.3915,0.4997)	0.3586 (0.2797,0.3858)	0.4295 (0.3535,0.4567)	0.4734 (0.3975,0.5007)	0.8607 (0.8335,0.8617)
OR	0.7528 (0.7067,0.7990)	0.7560 (0.7098,0.8022)	0.7519 (0.7057,0.7980)	0.7395 (0.6933,0.7857)	0.7727 (0.7265,0.8188)
RI	0.5455 (0.4147,0.6065)	0.5995 (0.4686,0.6605)	0.5360 (0.4062,0.5970)	0.7915 (0.6980,0.8525)	0.8622 (0.8012,0.8771)
SC	0.3103 (0.2550,0.3655)	0.3046 (0.2493,0.3599)	0.3062 (0.2509,0.3614)	0.2915 (0.2362,0.3467)	0.7603 (0.7051,0.8156)
SD	0.5021 (0.4144,0.5897)	0.3943 (0.3066,0.4820)	0.4407 (0.3531,0.5284)	0.4187 (0.3311,0.5064)	0.8576 (0.7700,0.8782)
TN	0.4679 (0.4333,0.5026)	0.4521 (0.4174,0.4867)	0.4328 (0.3981,0.4674)	0.4217 (0.3871,0.4563)	0.8589 (0.8242,0.8935)
TX	0.4782 (0.4646,0.4919)	0.5102 (0.4966,0.5239)	0.5308 (0.5172,0.5445)	0.5721 (0.5585,0.5858)	0.6869 (0.6732,0.7005)
UT	0.5835 (0.5095,0.6575)	0.5908 (0.5168,0.6649)	0.5794 (0.5054,0.6534)	0.5678 (0.4938,0.6418)	0.5617 (0.4877,0.6357)
VA	0.3282 (0.0687,0.5763)	0.3406 (0.1095,0.4770)	0.4008 (0.1589,0.5101)	0.4620 (0.2346,0.5321)	0.8226 (0.7828,0.8623)
VT	0.9304 (0.8744,0.9342)	0.9235 (0.8674,0.9275)	0.9174 (0.8614,0.9216)	0.9154 (0.8594,0.9196)	0.9169 (0.8609,0.9269)
WA	0.5768 (0.5539,0.5997)	0.5843 (0.5614,0.6072)	0.5743 (0.5514,0.5972)	0.5451 (0.5222,0.5680)	0.8861 (0.8632,0.9055)
WI	0.8636 (0.8285,0.8987)	0.8909 (0.8558,0.9259)	0.8844 (0.8493,0.9194)	0.8831 (0.8480,0.9182)	0.8468 (0.8117,0.8819)
WY	0.8509 (0.7413,0.8625)	0.8448 (0.7352,0.8565)	0.8322 (0.7227,0.8444)	0.8154 (0.7059,0.8285)	0.6929 (0.5833,0.7133)

Table 6b. Informed Coverage rates, July 1, 2009 – December 31, 2010, stratified by average income determined at the level of a 5-digit zip code: average income.

	Lowest Income	2nd Quartile	3rd Quartile	Highest Income	Missing
AK	-	-	0.4805 (0.0000,1.0000)	-	0.8512 (0.7878,0.9147)
AL	0.7433 (0.7205,0.7660)	0.7751 (0.7523,0.7978)	0.7589 (0.7361,0.7816)	0.7320 (0.7092,0.7548)	0.9289 (0.9062,0.9517)
AR	0.6908 (0.6596,0.7220)	0.6354 (0.6042,0.6666)	0.7441 (0.7129,0.7753)	0.7330 (0.7018,0.7642)	0.8950 (0.8638,0.9262)
AZ	-	-	0.2655 (0.0000,1.0000)	-	0.8964 (0.8808,0.9080)
CA	0.6437 (0.6348,0.6526)	0.6663 (0.6574,0.6752)	0.6662 (0.6573,0.6752)	0.6958 (0.6868,0.7047)	0.8432 (0.8342,0.8521)
CO	0.4531 (0.4078,0.4984)	0.5084 (0.4631,0.5537)	0.4868 (0.4414,0.5321)	0.4871 (0.4418,0.5325)	0.7347 (0.6894,0.7800)
CT	0.6153 (0.5862,0.6444)	0.6158 (0.5867,0.6449)	0.6525 (0.6234,0.6816)	0.6344 (0.6053,0.6634)	0.8993 (0.8702,0.9284)
DE	-	-	-	-	0.8333 (0.6612,0.9264)
FL	0.4022 (0.3824,0.4220)	0.4109 (0.3911,0.4307)	0.4062 (0.3864,0.4260)	0.4091 (0.3893,0.4289)	0.7257 (0.7059,0.7455)
GA	0.7425 (0.7079,1.0000)	0.5900 (0.3168,1.0000)	0.4765 (0.1706,0.8523)	0.7519 (0.5794,1.0000)	0.7638 (0.7292,0.7984)
HI	-	-	-	-	0.9000 (0.8070,0.9646)
IA	0.6191 (0.5682,0.6699)	0.7057 (0.6548,0.7565)	0.7220 (0.6711,0.7729)	0.6481 (0.5973,0.6990)	0.8455 (0.7946,0.8963)
ID	0.8328 (0.7838,0.8818)	0.8346 (0.7857,0.8836)	0.8380 (0.7890,0.8869)	0.8350 (0.7860,0.8840)	0.9043 (0.8553,0.9533)
IL	0.9580 (0.9480,0.9681)	0.9584 (0.9484,0.9685)	0.9602 (0.9502,0.9703)	0.9586 (0.9485,0.9687)	0.8868 (0.8768,0.8958)
IN	0.5416 (0.5078,0.5755)	0.5442 (0.5103,0.5780)	0.5249 (0.4910,0.5588)	0.5547 (0.5208,0.5885)	0.8757 (0.8419,0.9096)
KS	0.4612 (0.3551,0.5673)	0.4760 (0.3699,0.5821)	0.5562 (0.4502,0.6623)	0.4520 (0.3459,0.5581)	0.7502 (0.6441,0.8563)
LA	0.9498 (0.9325,0.9671)	0.9433 (0.9260,0.9606)	0.9390 (0.9217,0.9564)	0.9331 (0.9158,0.9504)	0.9375 (0.9202,0.9548)
MD	0.4886 (0.2371,0.6965)	0.8894 (0.8523,1.0000)	0.7676 (0.7378,0.8840)	0.8114 (0.7433,0.9598)	0.8694 (0.8396,0.8992)
MI	0.5720 (0.5440,0.6000)	0.5367 (0.5087,0.5647)	0.5348 (0.5068,0.5628)	0.5349 (0.5069,0.5629)	0.8510 (0.8230,0.8789)
MN	0.6291 (0.5864,0.6718)	0.7216 (0.6789,0.7643)	0.7497 (0.7071,0.7924)	0.6303 (0.5876,0.6729)	0.8066 (0.7639,0.8493)
MO	0.6448 (0.6181,0.6715)	0.6469 (0.6202,0.6736)	0.6658 (0.6390,0.6925)	0.6686 (0.6419,0.6953)	0.8868 (0.8601,0.9135)
MT	0.7342 (0.6449,0.8297)	0.7393 (0.6493,0.8347)	0.7375 (0.6477,0.8329)	0.7352 (0.6454,0.8306)	0.7810 (0.6897,0.8765)
NC	0.8106 (0.7846,0.8367)	0.8079 (0.7819,0.8339)	0.8013 (0.7752,0.8273)	0.7901 (0.7641,0.8161)	0.7989 (0.7729,0.8249)
ND	0.9227 (0.8370,0.9276)	0.9078 (0.8221,0.9134)	0.9050 (0.8192,0.9104)	0.9098 (0.8240,0.9155)	0.9038 (0.8180,0.9095)

	Lowest Income	2nd Quartile	3rd Quartile	Highest Income	Missing
NE	0.4279 (0.0485,0.6620)	0.7398 (0.6811,0.9606)	0.6879 (0.5066,0.9395)	0.5750 (0.5163,0.7268)	0.8400 (0.7814,0.8987)
NH	0.9224 (0.8524,0.9363)	0.9226 (0.8526,0.9366)	0.9162 (0.8463,0.9311)	0.9092 (0.8393,0.9246)	0.9062 (0.8363,0.9184)
NJ	0.8818 (0.8593,0.9044)	0.8813 (0.8588,0.9039)	0.8687 (0.8462,0.8913)	0.8641 (0.8416,0.8867)	0.8700 (0.8475,0.8926)
NM	0.9093 (0.8199,1.0000)	0.7885 (0.7652,0.8361)	0.9050 (0.8385,1.0000)	0.9500 (0.9267,1.0000)	0.9065 (0.8832,0.9298)
NV	-	0.1423 (0.0000,1.0000)	-	0.2242 (0.0000,0.9907)	0.4167 (0.3442,0.5038)
NY	0.9076 (0.8953,0.9199)	0.9031 (0.8908,0.9154)	0.8977 (0.8854,0.9100)	0.8897 (0.8774,0.9020)	0.9133 (0.9010,0.9256)
OK	0.5037 (0.4778,0.5295)	0.6610 (0.6352,0.6869)	0.5615 (0.5356,0.5873)	0.6068 (0.5809,0.6326)	0.8586 (0.8327,0.8844)
OR	0.7824 (0.7462,0.8187)	0.7830 (0.7467,0.8192)	0.7843 (0.7480,0.8206)	0.7749 (0.7386,0.8111)	0.8226 (0.7864,0.8589)
RI	0.7093 (0.5411,0.7512)	0.6920 (0.4657,0.7339)	0.7246 (0.5561,0.7665)	0.8054 (0.6682,0.8473)	0.9183 (0.8763,0.9199)
SC	0.3806 (0.3360,0.4422)	0.3767 (0.3323,0.4382)	0.4083 (0.3626,0.4699)	0.3661 (0.3222,0.4276)	0.7348 (0.6891,0.7964)
SD	0.7874 (0.7149,0.8599)	0.7854 (0.7129,0.8578)	0.7559 (0.6834,0.8284)	0.8127 (0.7402,0.8852)	0.9052 (0.8327,0.9256)
TN	0.6231 (0.5938,0.6524)	0.5696 (0.5403,0.5989)	0.5806 (0.5513,0.6099)	0.5170 (0.4877,0.5463)	0.8815 (0.8522,0.9108)
TX	0.5045 (0.4918,0.5172)	0.5197 (0.5070,0.5324)	0.5471 (0.5345,0.5598)	0.5923 (0.5796,0.6049)	0.7257 (0.7130,0.7384)
UT	0.6494 (0.5744,0.7245)	0.6572 (0.5822,0.7322)	0.6470 (0.5720,0.7220)	0.6448 (0.5697,0.7198)	0.6625 (0.5875,0.7376)
VA	0.6038 (0.5663,0.8327)	0.5238 (0.2653,0.8232)	0.3443 (0.2692,0.5384)	0.4170 (0.3794,0.4890)	0.8290 (0.7915,0.8666)
VT	0.9575 (0.9226,0.9606)	0.9516 (0.9149,0.9550)	0.9520 (0.9124,0.9553)	0.9449 (0.9023,0.9484)	0.9425 (0.9193,0.9470)
WA	0.5960 (0.5766,0.6154)	0.6431 (0.6237,0.6625)	0.6034 (0.5840,0.6228)	0.5733 (0.5539,0.5927)	0.9048 (0.8854,0.9242)
WI	0.5815 (0.5486,0.6144)	0.5796 (0.5467,0.6125)	0.6057 (0.5728,0.6386)	0.6043 (0.5715,0.6372)	0.8711 (0.8383,0.9040)
WY	0.8366 (0.7408,0.9100)	0.8320 (0.7363,0.9078)	0.8293 (0.7336,0.9042)	0.8153 (0.7195,0.8899)	0.8444 (0.7487,0.9172)

Table 7a. Informed Coverage rates, January 1, 2008 – June 30, 2009, stratified by percentage below the Federal Poverty Line (by quartile) determined at the level of a 5-digit zip code

	Lowest % below FPL	2nd Quartile	3rd Quartile	Highest % below FPL	Missing
AK	-	-	0.7756 (0.2016,1.0000)	-	0.8257 (0.7545,0.8372)
AL	0.5308 (0.4737,0.5619)	0.5219 (0.4636,0.5530)	0.4851 (0.4255,0.5163)	0.4953 (0.4360,0.5264)	0.8822 (0.8510,0.8926)
AR	0.4338 (0.3523,0.4666)	0.4057 (0.3271,0.4384)	0.5055 (0.4275,0.5383)	0.4735 (0.3936,0.5063)	0.8699 (0.8372,0.8709)
AZ	-	-	-	-	0.7612 (0.7391,0.7834)
CA	0.5140 (0.5036,0.5244)	0.4936 (0.4832,0.5040)	0.4953 (0.4849,0.5057)	0.4799 (0.4695,0.4903)	0.7997 (0.7893,0.8101)
CO	0.3616 (0.3085,0.4147)	0.3481 (0.2950,0.4012)	0.3608 (0.3078,0.4139)	0.3136 (0.2606,0.3667)	0.6634 (0.6104,0.7165)
CT	0.4848 (0.4298,0.5398)	0.4484 (0.3934,0.5034)	0.4726 (0.4176,0.5276)	0.4301 (0.3751,0.4851)	0.8012 (0.7463,0.8562)
DE	-	-	-	-	0.7778 (0.6174,0.8691)
FL	0.3457 (0.3205,0.3708)	0.3553 (0.3301,0.3804)	0.3633 (0.3381,0.3884)	0.3685 (0.3434,0.3936)	0.6196 (0.5945,0.6447)
GA	0.8645 (0.8259,0.9246)	0.8998 (0.7789,0.9687)	0.3840 (0.0000,0.4226)	0.2259 (0.0000,0.2936)	0.7474 (0.7088,0.7860)
HI	-	-	-	-	0.8571 (0.7412,0.9293)
IA	0.5408 (0.4369,0.5919)	0.5947 (0.4869,0.6458)	0.5812 (0.4788,0.6324)	0.5144 (0.4097,0.5655)	0.8723 (0.8211,0.8849)
ID	0.7673 (0.7085,0.8262)	0.7706 (0.7117,0.8295)	0.7744 (0.7155,0.8333)	0.7676 (0.7088,0.8265)	0.7900 (0.7311,0.8489)
IL	0.9243 (0.9126,0.9251)	0.9342 (0.9225,0.9350)	0.9427 (0.9310,0.9434)	0.9393 (0.9276,0.9400)	0.9252 (0.9135,0.9276)
IN	0.5804 (0.5417,0.6191)	0.5589 (0.5202,0.5976)	0.5685 (0.5298,0.6073)	0.5759 (0.5372,0.6147)	0.8423 (0.8035,0.8810)
KS	0.5056 (0.4394,0.5717)	0.5605 (0.4943,0.6266)	0.5310 (0.4649,0.5972)	0.5063 (0.4401,0.5724)	0.7329 (0.6668,0.7991)
LA	0.9149 (0.8934,0.9332)	0.9241 (0.9026,0.9406)	0.9339 (0.9124,0.9493)	0.9411 (0.9196,0.9566)	0.8163 (0.7948,0.8378)
MD	0.7329 (0.6608,0.7717)	0.8252 (0.7818,0.8640)	0.6778 (0.6193,0.7166)	0.4448 (0.2404,0.4836)	0.8306 (0.7918,0.8693)
MI	0.3113 (0.2608,0.3619)	0.2838 (0.2332,0.3343)	0.2909 (0.2404,0.3415)	0.2981 (0.2475,0.3486)	0.7001 (0.6495,0.7506)
MN	0.6170 (0.5715,0.6625)	0.6757 (0.6302,0.7212)	0.6697 (0.6243,0.7152)	0.5281 (0.4826,0.5736)	0.7252 (0.6797,0.7707)
MO	0.7369 (0.7065,0.7673)	0.7318 (0.7014,0.7623)	0.7409 (0.7105,0.7713)	0.7345 (0.7041,0.7649)	0.8509 (0.8205,0.8814)
MT	0.7039 (0.5989,0.8089)	0.7082 (0.6032,0.8132)	0.7194 (0.6144,0.8244)	0.7225 (0.6175,0.8275)	0.5307 (0.4257,0.6357)
NC	0.7534 (0.7249,0.7819)	0.7672 (0.7387,0.7957)	0.7750 (0.7465,0.8035)	0.7818 (0.7532,0.8103)	0.6712 (0.6427,0.6997)
ND	0.7448 (0.6179,0.8669)	0.7411 (0.6142,0.8632)	0.7357 (0.6088,0.8567)	0.7713 (0.6444,0.8797)	0.5702 (0.4433,0.6971)

	Lowest % below FPL	2nd Quartile	3rd Quartile	Highest % below FPL	Missing
NE	0.4104 (0.0612,0.4788)	0.5493 (0.1578,0.6177)	0.5376 (0.2776,0.6060)	0.4642 (0.1901,0.5347)	0.8334 (0.7650,0.8732)
NH	0.8596 (0.7761,0.8837)	0.8700 (0.7866,0.8928)	0.8734 (0.7899,0.8962)	0.8819 (0.7985,0.9039)	0.8613 (0.7778,0.8905)
NJ	0.8387 (0.8139,0.8636)	0.8459 (0.8210,0.8708)	0.8586 (0.8338,0.8835)	0.8742 (0.8493,0.8991)	0.7419 (0.7170,0.7668)
NM	0.7703 (0.6561,0.8396)	0.7345 (0.5607,0.8639)	0.8294 (0.6971,0.9432)	0.7797 (0.7023,0.8235)	0.8580 (0.8282,0.8877)
NV	0.2013 (0.0000,0.9010)	0.1067 (0.0000,0.4877)	-	-	0.3815 (0.2848,0.4781)
NY	0.8679 (0.8544,0.8689)	0.8858 (0.8722,0.8867)	0.9036 (0.8900,0.9044)	0.9013 (0.8878,0.9022)	0.7092 (0.6957,0.7154)
OK	0.4696 (0.3952,0.4968)	0.4187 (0.3425,0.4460)	0.4023 (0.3235,0.4295)	0.4387 (0.3546,0.4659)	0.8607 (0.8335,0.8617)
OR	0.7391 (0.6929,0.7853)	0.7505 (0.7043,0.7967)	0.7501 (0.7039,0.7962)	0.7606 (0.7144,0.8068)	0.7727 (0.7265,0.8188)
RI	0.6577 (0.5390,0.7187)	0.7369 (0.6273,0.7979)	0.4898 (0.3578,0.5508)	0.5615 (0.4294,0.6225)	0.8622 (0.8012,0.8771)
SC	0.2964 (0.2411,0.3517)	0.3070 (0.2518,0.3623)	0.2973 (0.2420,0.3526)	0.3108 (0.2555,0.3660)	0.7603 (0.7051,0.8156)
SD	0.4388 (0.3511,0.5264)	0.4091 (0.3214,0.4968)	0.4451 (0.3574,0.5328)	0.4676 (0.3799,0.5553)	0.8576 (0.7700,0.8782)
TN	0.4235 (0.3889,0.4582)	0.4533 (0.4186,0.4879)	0.4313 (0.3966,0.4659)	0.4656 (0.4310,0.5002)	0.8589 (0.8242,0.8935)
TX	0.5744 (0.5608,0.5881)	0.5510 (0.5374,0.5646)	0.4911 (0.4775,0.5048)	0.4751 (0.4614,0.4887)	0.6869 (0.6732,0.7005)
UT	0.5688 (0.4948,0.6428)	0.5794 (0.5054,0.6534)	0.5815 (0.5075,0.6555)	0.5918 (0.5178,0.6658)	0.5617 (0.4877,0.6357)
VA	0.3211 (0.0162,0.4545)	0.3759 (0.1501,0.5249)	0.4664 (0.2967,0.5062)	0.4011 (0.1581,0.6148)	0.8226 (0.7828,0.8623)
VT	0.9141 (0.8580,0.9183)	0.9182 (0.8621,0.9222)	0.9226 (0.8665,0.9267)	0.9319 (0.8759,0.9357)	0.9169 (0.8609,0.9269)
WA	0.5537 (0.5308,0.5767)	0.5872 (0.5643,0.6101)	0.5758 (0.5529,0.5987)	0.5639 (0.5410,0.5868)	0.8861 (0.8632,0.9055)
WI	0.8776 (0.8425,0.9126)	0.8917 (0.8566,0.9267)	0.8960 (0.8610,0.9311)	0.8569 (0.8219,0.8920)	0.8468 (0.8117,0.8819)
WY	0.8234 (0.7139,0.8360)	0.8251 (0.7156,0.8376)	0.8397 (0.7302,0.8519)	0.8556 (0.7461,0.8670)	0.6929 (0.5833,0.7133)

Table 7b. Informed Coverage rates, July 1, 2009 – December 31, 2010, stratified by percentage below the Federal Poverty Line (by quartile) determined at the level of a 5-digit zip code

	Lowest % below FPL	2nd Quartile	3rd Quartile	Highest % below FPL	Missing
AK	-	-	0.4805 (0.0000,1.0000)	-	0.8512 (0.7878,0.9147)
AL	0.7943 (0.7716,0.8171)	0.7246 (0.7019,0.7474)	0.7294 (0.7067,0.7522)	0.7604 (0.7376,0.7831)	0.9289 (0.9062,0.9517)
AR	0.7436 (0.7124,0.7748)	0.7065 (0.6753,0.7377)	0.6512 (0.6200,0.6824)	0.6938 (0.6626,0.7250)	0.8950 (0.8638,0.9262)
AZ	-	-	0.2655 (0.0000,1.0000)	-	0.8964 (0.8808,0.9080)
CA	0.7078 (0.6988,0.7167)	0.6601 (0.6512,0.6691)	0.6597 (0.6508,0.6686)	0.6429 (0.6340,0.6519)	0.8432 (0.8342,0.8521)
CO	0.4848 (0.4395,0.5301)	0.4747 (0.4294,0.5200)	0.5491 (0.5038,0.5944)	0.4143 (0.3690,0.4596)	0.7347 (0.6894,0.7800)
CT	0.6449 (0.6158,0.6740)	0.6598 (0.6307,0.6889)	0.6033 (0.5742,0.6323)	0.6074 (0.5783,0.6365)	0.8993 (0.8702,0.9284)
DE	-	-	-	-	0.8333 (0.6612,0.9264)
FL	0.4129 (0.3931,0.4327)	0.4114 (0.3916,0.4312)	0.4071 (0.3873,0.4269)	0.3970 (0.3772,0.4168)	0.7257 (0.7059,0.7455)
GA	0.8192 (0.6814,1.0000)	0.4615 (0.1512,0.8380)	0.5900 (0.3168,1.0000)	0.6218 (0.5873,0.9049)	0.7638 (0.7292,0.7984)
HI	-	-	-	-	0.9000 (0.8070,0.9646)
IA	0.7239 (0.6730,0.7748)	0.6903 (0.6394,0.7411)	0.6775 (0.6266,0.7284)	0.6001 (0.5492,0.6510)	0.8455 (0.7946,0.8963)
ID	0.8356 (0.7867,0.8846)	0.8342 (0.7853,0.8832)	0.8368 (0.7878,0.8858)	0.8338 (0.7848,0.8827)	0.9043 (0.8553,0.9533)
IL	0.9550 (0.9449,0.9651)	0.9572 (0.9471,0.9672)	0.9615 (0.9514,0.9716)	0.9616 (0.9515,0.9717)	0.8868 (0.8768,0.8958)
IN	0.5590 (0.5251,0.5929)	0.5398 (0.5059,0.5736)	0.5505 (0.5166,0.5844)	0.5122 (0.4784,0.5461)	0.8757 (0.8419,0.9096)
KS	0.5521 (0.4460,0.6581)	0.4839 (0.3778,0.5900)	0.4869 (0.3808,0.5929)	0.4203 (0.3143,0.5264)	0.7502 (0.6441,0.8563)
LA	0.9327 (0.9154,0.9500)	0.9373 (0.9200,0.9546)	0.9449 (0.9276,0.9622)	0.9505 (0.9332,0.9678)	0.9375 (0.9202,0.9548)
MD	0.8562 (0.8264,0.9761)	0.7159 (0.6860,0.8588)	0.8267 (0.7621,0.9632)	0.6000 (0.3785,0.8143)	0.8694 (0.8396,0.8992)
MI	0.5197 (0.4917,0.5477)	0.5738 (0.5458,0.6017)	0.4954 (0.4674,0.5234)	0.5793 (0.5513,0.6073)	0.8510 (0.8230,0.8789)
MN	0.6857 (0.6431,0.7284)	0.7607 (0.7180,0.8034)	0.7375 (0.6949,0.7802)	0.5379 (0.4953,0.5806)	0.8066 (0.7639,0.8493)
MO	0.6713 (0.6446,0.6980)	0.6670 (0.6403,0.6937)	0.6466 (0.6199,0.6733)	0.6407 (0.6140,0.6674)	0.8868 (0.8601,0.9135)
MT	0.7338 (0.6440,0.8293)	0.7386 (0.6487,0.8340)	0.7399 (0.6499,0.8353)	0.7340 (0.6447,0.8294)	0.7810 (0.6897,0.8765)
NC	0.7918 (0.7658,0.8178)	0.8027 (0.7767,0.8288)	0.8061 (0.7801,0.8322)	0.8092 (0.7832,0.8353)	0.7989 (0.7729,0.8249)

	Lowest % below FPL	2nd Quartile	3rd Quartile	Highest % below FPL	Missing
ND	0.9082 (0.8225,0.9137)	0.9097 (0.8239,0.9150)	0.9007 (0.8150,0.9067)	0.9261 (0.8403,0.9310)	0.9038 (0.8180,0.9095)
NE	0.5750 (0.5163,0.7268)	0.6699 (0.4495,0.9515)	0.6438 (0.5067,0.8791)	0.6204 (0.5617,0.8214)	0.8400 (0.7814,0.8987)
NH	0.9100 (0.8401,0.9252)	0.9178 (0.8479,0.9326)	0.9193 (0.8493,0.9337)	0.9236 (0.8536,0.9374)	0.9062 (0.8363,0.9184)
NJ	0.8618 (0.8392,0.8843)	0.8706 (0.8480,0.8931)	0.8766 (0.8540,0.8992)	0.8869 (0.8643,0.9094)	0.8700 (0.8475,0.8926)
NM	0.8952 (0.8085,1.0000)	0.8725 (0.8491,0.9730)	0.8739 (0.8505,0.9941)	0.9263 (0.8815,1.0000)	0.9065 (0.8832,0.9298)
NV	-	0.1423 (0.0000,1.0000)	0.2242 (0.0000,0.9907)	-	0.4167 (0.3442,0.5038)
NY	0.8835 (0.8712,0.8958)	0.8977 (0.8853,0.9100)	0.9087 (0.8964,0.9210)	0.9079 (0.8956,0.9202)	0.9133 (0.9010,0.9256)
OK	0.6166 (0.5907,0.6424)	0.5762 (0.5503,0.6020)	0.5845 (0.5586,0.6103)	0.5633 (0.5374,0.5891)	0.8586 (0.8327,0.8844)
OR	0.7737 (0.7374,0.8099)	0.7824 (0.7461,0.8186)	0.7807 (0.7444,0.8169)	0.7879 (0.7517,0.8242)	0.8226 (0.7864,0.8589)
RI	0.7627 (0.6105,0.8046)	0.7209 (0.5604,0.7628)	0.7729 (0.5738,0.8149)	0.7176 (0.5446,0.7595)	0.9183 (0.8763,0.9199)
SC	0.3863 (0.3415,0.4479)	0.3973 (0.3523,0.4589)	0.3769 (0.3324,0.4385)	0.3714 (0.3272,0.4330)	0.7348 (0.6891,0.7964)
SD	0.8123 (0.7398,0.8848)	0.7585 (0.6860,0.8310)	0.7926 (0.7201,0.8651)	0.7824 (0.7099,0.8549)	0.9052 (0.8327,0.9256)
TN	0.4952 (0.4659,0.5245)	0.5828 (0.5535,0.6121)	0.6080 (0.5787,0.6373)	0.6059 (0.5766,0.6351)	0.8815 (0.8522,0.9108)
TX	0.5947 (0.5820,0.6074)	0.5535 (0.5408,0.5662)	0.5238 (0.5111,0.5364)	0.4914 (0.4788,0.5041)	0.7257 (0.7130,0.7384)
UT	0.6436 (0.5686,0.7186)	0.6495 (0.5745,0.7245)	0.6495 (0.5745,0.7245)	0.6557 (0.5807,0.7308)	0.6625 (0.5875,0.7376)
VA	0.4712 (0.4337,0.6205)	0.6003 (0.5628,0.8196)	0.2769 (0.0000,0.5428)	0.5594 (0.5083,0.7883)	0.8290 (0.7915,0.8666)
VT	0.9460 (0.9035,0.9496)	0.9489 (0.9097,0.9522)	0.9525 (0.9135,0.9559)	0.9586 (0.9253,0.9617)	0.9425 (0.9193,0.9470)
WA	0.5785 (0.5590,0.5979)	0.6296 (0.6102,0.6491)	0.6055 (0.5861,0.6249)	0.6061 (0.5867,0.6255)	0.9048 (0.8854,0.9242)
WI	0.6105 (0.5776,0.6433)	0.6017 (0.5688,0.6346)	0.5847 (0.5519,0.6176)	0.5738 (0.5409,0.6066)	0.8711 (0.8383,0.9040)
WY	0.8262 (0.7304,0.9024)	0.8173 (0.7216,0.8919)	0.8313 (0.7356,0.9059)	0.8388 (0.7431,0.9121)	0.8444 (0.7487,0.9172)

Table 8a. Informed Coverage rates, June 1, 2008 – June 30, 2009, stratified by percentage with High School Degree (by quartile) determined at the level of a 5-digit zip code

	Lowest % with degree	2nd Quartile	3rd Quartile	Highest % with degree	Missing
AK	-	-	0.7756 (0.2016,1.0000)	-	0.8257 (0.7545,0.8372)
AL	0.4867 (0.4295,0.5178)	0.4994 (0.4404,0.5305)	0.5164 (0.4547,0.5475)	0.5307 (0.4737,0.5618)	0.8822 (0.8510,0.8926)
AR	0.4915 (0.4143,0.5242)	0.4866 (0.4000,0.5194)	0.4107 (0.3342,0.4435)	0.4390 (0.3601,0.4718)	0.8699 (0.8372,0.8709)
AZ	-	-	-	-	0.7612 (0.7391,0.7834)
CA	0.4729 (0.4625,0.4833)	0.4804 (0.4700,0.4908)	0.5058 (0.4954,0.5162)	0.5248 (0.5144,0.5352)	0.7997 (0.7893,0.8101)
CO	0.3153 (0.2622,0.3684)	0.3713 (0.3182,0.4244)	0.3508 (0.2978,0.4039)	0.3458 (0.2927,0.3988)	0.6634 (0.6104,0.7165)
CT	0.4274 (0.3724,0.4824)	0.4650 (0.4100,0.5200)	0.4661 (0.4111,0.5210)	0.4781 (0.4231,0.5330)	0.8012 (0.7463,0.8562)
DE	-	-	-	-	0.7778 (0.6174,0.8691)
FL	0.3639 (0.3387,0.3890)	0.3626 (0.3375,0.3878)	0.3538 (0.3287,0.3789)	0.3526 (0.3275,0.3778)	0.6196 (0.5945,0.6447)
GA	0.2259 (0.0000,0.2936)	0.7826 (0.6571,0.8805)	0.7112 (0.4857,0.8501)	0.8178 (0.7792,0.8564)	0.7474 (0.7088,0.7860)
HI	-	-	-	-	0.8571 (0.7412,0.9293)
IA	0.4776 (0.3697,0.5287)	0.5652 (0.4600,0.6164)	0.6196 (0.5142,0.6707)	0.5683 (0.4692,0.6195)	0.8723 (0.8211,0.8849)
ID	0.7806 (0.7217,0.8394)	0.7712 (0.7123,0.8301)	0.7710 (0.7121,0.8299)	0.7578 (0.6989,0.8166)	0.7900 (0.7311,0.8489)
IL	0.9516 (0.9399,0.9523)	0.9319 (0.9202,0.9327)	0.9333 (0.9217,0.9341)	0.9230 (0.9114,0.9239)	0.9252 (0.9135,0.9276)
IN	0.5695 (0.5308,0.6083)	0.5669 (0.5282,0.6056)	0.5795 (0.5408,0.6182)	0.5674 (0.5286,0.6061)	0.8423 (0.8035,0.8810)
KS	0.5345 (0.4684,0.6007)	0.5103 (0.4441,0.5764)	0.5278 (0.4616,0.5939)	0.5311 (0.4649,0.5972)	0.7329 (0.6668,0.7991)
LA	0.9371 (0.9157,0.9523)	0.9346 (0.9131,0.9498)	0.9247 (0.9032,0.9422)	0.9178 (0.8963,0.9357)	0.8163 (0.7948,0.8378)
MD	0.4656 (0.2590,0.5059)	0.7682 (0.7248,0.8069)	0.8019 (0.7631,0.8406)	0.7258 (0.6826,0.7645)	0.8306 (0.7918,0.8693)
MI	0.2900 (0.2394,0.3405)	0.3032 (0.2527,0.3538)	0.2637 (0.2132,0.3143)	0.3263 (0.2758,0.3769)	0.7001 (0.6495,0.7506)
MN	0.5822 (0.5367,0.6276)	0.6814 (0.6359,0.7268)	0.6833 (0.6378,0.7288)	0.5108 (0.4653,0.5563)	0.7252 (0.6797,0.7707)
MO	0.7253 (0.6949,0.7557)	0.7316 (0.7011,0.7620)	0.7352 (0.7048,0.7657)	0.7521 (0.7216,0.7825)	0.8509 (0.8205,0.8814)
MT	0.7154 (0.6104,0.8204)	0.7178 (0.6128,0.8228)	0.7121 (0.6071,0.8170)	0.7079 (0.6029,0.8129)	0.5307 (0.4257,0.6357)
NC	0.7863 (0.7577,0.8148)	0.7794 (0.7509,0.8080)	0.7662 (0.7377,0.7947)	0.7451 (0.7166,0.7736)	0.6712 (0.6427,0.6997)
ND	0.7653 (0.6384,0.8787)	0.7360 (0.6091,0.8583)	0.7467 (0.6198,0.8657)	0.7465 (0.6196,0.8654)	0.5702 (0.4433,0.6971)

	Lowest % with degree	2nd Quartile	3rd Quartile	Highest % with degree	Missing
NE	0.5633 (0.3182,0.6317)	0.4941 (0.1846,0.5625)	0.4886 (0.0722,0.5570)	0.3962 (0.0747,0.4646)	0.8334 (0.7650,0.8732)
NH	0.8813 (0.7979,0.9033)	0.8788 (0.7954,0.9007)	0.8671 (0.7836,0.8901)	0.8580 (0.7746,0.8827)	0.8613 (0.7778,0.8905)
NJ	0.8705 (0.8456,0.8954)	0.8575 (0.8326,0.8824)	0.8518 (0.8269,0.8767)	0.8385 (0.8137,0.8634)	0.7419 (0.7170,0.7668)
NM	0.6444 (0.4404,0.7443)	0.8204 (0.7108,0.9152)	0.7805 (0.6023,0.9326)	0.8240 (0.7529,0.8854)	0.8580 (0.8282,0.8877)
NV	-	-	0.3294 (0.0000,1.0000)	0.0317 (0.0000,0.2552)	0.3815 (0.2848,0.4781)
NY	0.9023 (0.8888,0.9032)	0.8964 (0.8829,0.8973)	0.8866 (0.8731,0.8875)	0.8742 (0.8607,0.8752)	0.7093 (0.6957,0.7155)
OK	0.4563 (0.3683,0.4836)	0.4163 (0.3358,0.4435)	0.4202 (0.3481,0.4474)	0.4413 (0.3667,0.4686)	0.8607 (0.8335,0.8617)
OR	0.7617 (0.7156,0.8079)	0.7521 (0.7059,0.7983)	0.7493 (0.7031,0.7955)	0.7370 (0.6908,0.7832)	0.7727 (0.7265,0.8188)
RI	0.4884 (0.3529,0.5494)	0.6357 (0.5156,0.6967)	0.7164 (0.6066,0.7774)	0.6116 (0.4827,0.6726)	0.8622 (0.8012,0.8771)
SC	0.3140 (0.2587,0.3692)	0.2984 (0.2431,0.3536)	0.2979 (0.2427,0.3532)	0.3016 (0.2463,0.3568)	0.7603 (0.7051,0.8156)
SD	0.4796 (0.3919,0.5673)	0.4380 (0.3503,0.5257)	0.3988 (0.3111,0.4865)	0.4412 (0.3535,0.5289)	0.8576 (0.7700,0.8782)
TN	0.4525 (0.4179,0.4872)	0.4620 (0.4274,0.4967)	0.4200 (0.3854,0.4547)	0.4395 (0.4048,0.4741)	0.8589 (0.8242,0.8935)
TX	0.4914 (0.4777,0.5050)	0.5055 (0.4918,0.5191)	0.5192 (0.5056,0.5329)	0.5758 (0.5621,0.5894)	0.6869 (0.6732,0.7005)
UT	0.5851 (0.5111,0.6591)	0.5821 (0.5081,0.6561)	0.5721 (0.4981,0.6461)	0.5829 (0.5089,0.6569)	0.5617 (0.4877,0.6357)
VA	0.5832 (0.4258,0.6917)	0.3378 (0.0942,0.5545)	0.4802 (0.2521,0.5725)	0.2050 (0.0000,0.3089)	0.8226 (0.7828,0.8623)
VT	0.9279 (0.8718,0.9317)	0.9222 (0.8661,0.9262)	0.9188 (0.8628,0.9229)	0.9178 (0.8618,0.9220)	0.9169 (0.8609,0.9269)
WA	0.5648 (0.5419,0.5877)	0.5739 (0.5510,0.5968)	0.5841 (0.5612,0.6070)	0.5588 (0.5359,0.5818)	0.8861 (0.8632,0.9055)
WI	0.8718 (0.8367,0.9069)	0.8745 (0.8394,0.9096)	0.8908 (0.8557,0.9258)	0.8855 (0.8505,0.9206)	0.8468 (0.8117,0.8819)
WY	0.8519 (0.7423,0.8635)	0.8378 (0.7282,0.8497)	0.8215 (0.7120,0.8345)	0.8324 (0.7229,0.8445)	0.6930 (0.5835,0.7134)

Table 8b. Informed Coverage rates, July 1, 2009 – December 31, 2010, stratified by percentage with High School Degree (by quartile) determined at the level of a 5-digit zip code

	Lowest % with degree	2nd Quartile	3rd Quartile	Highest % with degree	Missing
AK	-	-	0.4805 (0.0000,1.0000)	-	0.8512 (0.7878,0.9147)
AL	0.7513 (0.7286,0.7741)	0.7505 (0.7277,0.7732)	0.7220 (0.6993,0.7448)	0.7865 (0.7638,0.8093)	0.9289 (0.9062,0.9517)
AR	0.6926 (0.6614,0.7238)	0.6735 (0.6422,0.7047)	0.7240 (0.6928,0.7552)	0.7076 (0.6764,0.7388)	0.8950 (0.8638,0.9262)
AZ	-	-	0.2655 (0.0000,1.0000)	-	0.8964 (0.8808,0.9080)
CA	0.6515 (0.6426,0.6604)	0.6449 (0.6360,0.6538)	0.6807 (0.6717,0.6896)	0.6948 (0.6859,0.7037)	0.8432 (0.8342,0.8521)
CO	0.4695 (0.4242,0.5148)	0.4945 (0.4492,0.5398)	0.5145 (0.4692,0.5598)	0.4515 (0.4062,0.4969)	0.7347 (0.6894,0.7800)
CT	0.6198 (0.5907,0.6488)	0.6194 (0.5903,0.6484)	0.6466 (0.6175,0.6757)	0.6360 (0.6069,0.6651)	0.8993 (0.8702,0.9284)
DE	-	-	-	-	0.8333 (0.6612,0.9264)
FL	0.3869 (0.3671,0.4067)	0.4239 (0.4041,0.4438)	0.4152 (0.3954,0.4350)	0.4015 (0.3817,0.4213)	0.7257 (0.7059,0.7455)
GA	0.7425 (0.7079,1.0000)	0.4864 (0.0000,0.9227)	0.4765 (0.1706,0.8523)	0.8437 (0.7625,0.9798)	0.7638 (0.7292,0.7984)
HI	-	-	-	-	0.9000 (0.8070,0.9646)
IA	0.6154 (0.5645,0.6663)	0.6797 (0.6288,0.7305)	0.6965 (0.6457,0.7474)	0.6996 (0.6487,0.7505)	0.8455 (0.7946,0.8963)
ID	0.8437 (0.7947,0.8926)	0.8333 (0.7843,0.8822)	0.8348 (0.7859,0.8838)	0.8289 (0.7799,0.8779)	0.9043 (0.8553,0.9533)
IL	0.9677 (0.9576,0.9777)	0.9587 (0.9486,0.9687)	0.9558 (0.9457,0.9658)	0.9529 (0.9428,0.9629)	0.8868 (0.8768,0.8958)
IN	0.5207 (0.4868,0.5546)	0.5399 (0.5061,0.5738)	0.5690 (0.5352,0.6029)	0.5330 (0.4991,0.5668)	0.8757 (0.8419,0.9096)
KS	0.4467 (0.3406,0.5528)	0.5623 (0.4562,0.6684)	0.4252 (0.3192,0.5313)	0.4981 (0.3920,0.6041)	0.7502 (0.6441,0.8563)
LA	0.9472 (0.9299,0.9646)	0.9450 (0.9277,0.9623)	0.9393 (0.9220,0.9566)	0.9338 (0.9165,0.9511)	0.9375 (0.9202,0.9548)
MD	0.5223 (0.2779,0.7307)	0.8211 (0.7913,0.9436)	0.7891 (0.6786,0.9546)	0.8496 (0.8197,0.9723)	0.8694 (0.8396,0.8992)
MI	0.5656 (0.5376,0.5936)	0.5341 (0.5061,0.5621)	0.5601 (0.5321,0.5881)	0.5178 (0.4898,0.5458)	0.8510 (0.8230,0.8789)
MN	0.6127 (0.5700,0.6554)	0.7841 (0.7415,0.8268)	0.7482 (0.7055,0.7909)	0.6048 (0.5621,0.6475)	0.8066 (0.7639,0.8493)
MO	0.6491 (0.6224,0.6758)	0.6582 (0.6315,0.6849)	0.6571 (0.6304,0.6838)	0.6616 (0.6349,0.6883)	0.8868 (0.8601,0.9135)
MT	0.7305 (0.6414,0.8259)	0.7382 (0.6483,0.8336)	0.7407 (0.6506,0.8361)	0.7362 (0.6464,0.8316)	0.7810 (0.6897,0.8765)
NC	0.8135 (0.7874,0.8395)	0.8094 (0.7834,0.8354)	0.8000 (0.7740,0.8260)	0.7867 (0.7606,0.8127)	0.7989 (0.7729,0.8249)
ND	0.9240 (0.8383,0.9290)	0.9034 (0.8177,0.9090)	0.9082 (0.8224,0.9137)	0.9102 (0.8245,0.9157)	0.9038 (0.8180,0.9095)

	Lowest % with degree	2nd Quartile	3rd Quartile	Highest % with degree	Missing
NE	0.6328 (0.5049,0.8775)	0.7019 (0.6433,0.9373)	0.6186 (0.3895,0.8907)	0.5750 (0.5163,0.7268)	0.8400 (0.7814,0.8987)
NH	0.9222 (0.8523,0.9361)	0.9226 (0.8526,0.9367)	0.9160 (0.8461,0.9308)	0.9099 (0.8400,0.9253)	0.9062 (0.8363,0.9184)
NJ	0.8838 (0.8612,0.9064)	0.8760 (0.8534,0.8986)	0.8746 (0.8520,0.8971)	0.8620 (0.8395,0.8846)	0.8700 (0.8475,0.8926)
NM	0.8952 (0.8573,1.0000)	0.9312 (0.8892,1.0000)	0.8100 (0.7867,0.8646)	0.9158 (0.8494,1.0000)	0.9065 (0.8832,0.9298)
NV	-	0.2242 (0.0000,0.9907)	-	0.1423 (0.0000,1.0000)	0.4167 (0.3442,0.5038)
NY	0.9088 (0.8965,0.9211)	0.9041 (0.8918,0.9164)	0.8975 (0.8851,0.9098)	0.8877 (0.8754,0.9000)	0.9133 (0.9010,0.9256)
OK	0.5856 (0.5598,0.6115)	0.5734 (0.5475,0.5992)	0.5682 (0.5423,0.5940)	0.6118 (0.5859,0.6376)	0.8586 (0.8327,0.8844)
OR	0.7875 (0.7512,0.8238)	0.7846 (0.7484,0.8209)	0.7799 (0.7437,0.8162)	0.7725 (0.7363,0.8088)	0.8226 (0.7864,0.8589)
RI	0.7769 (0.5988,0.8188)	0.7341 (0.5643,0.7761)	0.6467 (0.4630,0.6886)	0.8072 (0.6660,0.8491)	0.9183 (0.8763,0.9199)
SC	0.3746 (0.3303,0.4362)	0.4126 (0.3669,0.4742)	0.3555 (0.3120,0.4171)	0.3858 (0.3410,0.4474)	0.7348 (0.6891,0.7964)
SD	0.7718 (0.6993,0.8443)	0.8086 (0.7361,0.8811)	0.7939 (0.7214,0.8664)	0.7748 (0.7024,0.8473)	0.9052 (0.8327,0.9256)
TN	0.6184 (0.5892,0.6477)	0.5949 (0.5656,0.6242)	0.5367 (0.5074,0.5660)	0.5469 (0.5176,0.5762)	0.8815 (0.8522,0.9108)
TX	0.5063 (0.4936,0.5189)	0.5080 (0.4953,0.5207)	0.5529 (0.5403,0.5656)	0.5987 (0.5860,0.6113)	0.7257 (0.7130,0.7384)
UT	0.6528 (0.5778,0.7278)	0.6484 (0.5734,0.7234)	0.6466 (0.5716,0.7217)	0.6508 (0.5758,0.7258)	0.6625 (0.5875,0.7376)
VA	0.3054 (0.0000,0.6048)	0.6188 (0.5351,0.8489)	0.5798 (0.5422,0.7862)	0.3878 (0.3502,0.4451)	0.8290 (0.7915,0.8666)
VT	0.9558 (0.9196,0.9589)	0.9517 (0.9142,0.9551)	0.9512 (0.9134,0.9546)	0.9473 (0.9049,0.9508)	0.9425 (0.9193,0.9470)
WA	0.5850 (0.5656,0.6044)	0.6056 (0.5862,0.6250)	0.6300 (0.6106,0.6494)	0.5990 (0.5796,0.6185)	0.9048 (0.8854,0.9242)
WI	0.5798 (0.5469,0.6127)	0.5841 (0.5512,0.6169)	0.5940 (0.5612,0.6269)	0.6127 (0.5798,0.6456)	0.8711 (0.8383,0.9040)
WY	0.8364 (0.7406,0.9093)	0.8311 (0.7354,0.9053)	0.8180 (0.7223,0.8939)	0.8272 (0.7314,0.9029)	0.8444 (0.7487,0.9173)

Table 9a. Informed Coverage rates stratified by geographic category, January 1, 2008 – June 30, 2009

	Rural	Urban Cluster	Urbanized Area	Missing Geography
AK	-	-	0.7756 (1.0000,0.2016)	0.8257 (0.6615,0.7545)
AL	0.4978 (0.4133,0.5289)	0.4489 (0.3833,0.4800)	0.5303 (0.2846,0.4992)	0.8822 (0.7227,0.8511)
AR	0.7778 (0.7425,0.8105)	0.4279 (0.3503,0.4607)	0.4921 (0.2497,0.4418)	0.8700 (0.7325,0.8372)
AZ	-	-	-	0.7612 (0.6150,0.7391)
CA	0.5145 (0.5041,0.5249)	0.5089 (0.4985,0.5193)	0.4907 (0.2665,0.4803)	0.7998 (0.6106,0.7894)
CO	0.3570 (0.3039,0.4101)	0.3351 (0.2821,0.3882)	0.3513 (0.2806,0.2982)	0.6635 (0.5867,0.6104)
CT	0.4577 (0.4027,0.5127)	0.4475 (0.3925,0.5025)	0.4594 (0.3337,0.4044)	0.8013 (0.7295,0.7463)
DE	-	-	-	0.7778 (0.7733,0.6174)
FL	0.3681 (0.3429,0.3932)	0.3552 (0.3301,0.3804)	0.3571 (0.2660,0.3320)	0.6196 (0.5248,0.5945)
GA	0.7041 (0.4848,0.7595)	-	0.6780 (0.6427,0.5905)	0.7474 (0.5816,0.7088)
HI	-	-	-	0.8571 (0.8243,0.7412)
IA	0.7460 (0.6469,0.7972)	0.6352 (0.5274,0.6863)	0.5313 (0.3693,0.4724)	0.8723 (0.6817,0.8211)
ID	0.7649 (0.7060,0.8237)	0.7704 (0.7115,0.8292)	0.7720 (0.6782,0.7131)	0.7077 (0.7483,0.6488)
IL	0.9212 (0.9095,0.9229)	0.9241 (0.9124,0.9253)	0.9374 (0.8174,0.9257)	0.9529 (0.8558,0.9259)
IN	0.6156 (0.5768,0.6543)	0.5826 (0.5439,0.6214)	0.5612 (0.3841,0.5225)	0.8423 (0.7329,0.8036)
KS	0.5648 (0.4986,0.6309)	0.5178 (0.4517,0.5840)	0.5151 (0.3510,0.4490)	0.7330 (0.6015,0.6668)
LA	0.9262 (0.9047,0.9419)	0.9324 (0.9109,0.9468)	0.9284 (0.7756,0.9069)	0.5191 (0.1997,0.4976)
MD	0.5753 (0.0160,0.6141)	0.7804 (0.7416,0.8191)	0.6464 (0.5998,0.5991)	0.8306 (0.7118,0.7918)
MI	0.3207 (0.2701,0.3713)	0.2590 (0.2085,0.3096)	0.3019 (0.2677,0.2513)	0.7001 (0.6857,0.6495)
MN	0.4986 (0.4531,0.5440)	0.4633 (0.4179,0.5088)	0.4603 (0.3183,0.4148)	0.7251 (0.6063,0.6796)
MO	0.7829 (0.7525,0.8133)	0.7313 (0.7008,0.7617)	0.7363 (0.6476,0.7058)	0.8509 (0.7205,0.8205)
MT	0.7160 (0.6110,0.8210)	0.7067 (0.6017,0.8117)	0.7132 (0.6502,0.6082)	0.4540 (0.2979,0.3490)
NC	0.7802 (0.7516,0.8087)	0.7846 (0.7561,0.8131)	0.7615 (0.6825,0.7329)	0.4579 (0.3513,0.4294)

	Rural	Urban Cluster	Urbanized Area	Missing Geography
ND	0.7524 (0.6255,0.8693)	0.7421 (0.6152,0.8637)	0.7493 (0.5902,0.6224)	0.3691 (0.2195,0.2422)
NE	0.5436 (0.1002,0.7410)	0.6824 (0.3620,0.7508)	0.4274 (0.2706,0.2722)	0.8334 (0.6860,0.7650)
NH	0.8715 (0.7880,0.8944)	0.8770 (0.7935,0.8990)	0.8688 (0.7663,0.7853)	0.7893 (0.6434,0.7059)
NJ	0.8341 (0.8093,0.8590)	0.8245 (0.7997,0.8494)	0.8556 (0.6912,0.8307)	0.5332 (0.2764,0.5083)
NM	0.7421 (0.5991,0.8259)	0.6811 (0.4067,0.9341)	0.8035 (0.7897,0.7705)	0.8580 (0.7500,0.8282)
NV	-	-	0.3938 (0.6484,0.2971)	0.3815 (0.3788,0.2848)
NY	0.8705 (0.8569,0.8723)	0.8770 (0.8635,0.8790)	0.8923 (0.7009,0.8787)	0.5982 (0.3199,0.5847)
OK	0.4118 (0.3163,0.4390)	0.4279 (0.3550,0.4552)	0.4480 (0.2030,0.3963)	0.8607 (0.6835,0.8335)
OR	0.7504 (0.7042,0.7966)	0.7522 (0.7061,0.7984)	0.7498 (0.6170,0.7037)	0.8115 (0.7183,0.7653)
RI	0.7121 (0.4940,0.7801)	0.7702 (0.6335,0.8312)	0.5877 (0.4675,0.5232)	0.8622 (0.7480,0.8012)
SC	0.3241 (0.2688,0.3793)	0.3035 (0.2482,0.3587)	0.3016 (0.2173,0.2463)	0.7604 (0.6930,0.7051)
SD	0.4882 (0.4005,0.5759)	0.4218 (0.3341,0.5094)	0.4319 (0.2520,0.3442)	0.8577 (0.7333,0.7700)
TN	0.4337 (0.3990,0.4683)	0.4398 (0.4051,0.4744)	0.4447 (0.2840,0.4101)	0.8589 (0.7446,0.8243)
TX	0.5210 (0.5074,0.5346)	0.5550 (0.5414,0.5687)	0.5110 (0.3697,0.4973)	0.6869 (0.5368,0.6732)
UT	0.5848 (0.5108,0.6588)	0.5718 (0.4978,0.6458)	0.5841 (0.4917,0.5101)	0.4814 (0.3907,0.4074)
VA	0.2141 (0.0000,0.4009)	0.4233 (0.2157,0.5793)	0.4780 (0.4600,0.3714)	0.8226 (0.7278,0.7828)
VT	0.9182 (0.8622,0.9212)	0.9252 (0.8692,0.9284)	0.9270 (0.8476,0.8709)	0.8477 (0.6824,0.7916)
WA	0.5894 (0.5665,0.6123)	0.5744 (0.5515,0.5974)	0.5612 (0.3335,0.5383)	0.8862 (0.7577,0.8633)
WI	0.8727 (0.8377,0.9078)	0.8788 (0.8437,0.9139)	0.8844 (0.8494,0.8493)	0.8468 (0.7313,0.8117)
WY	0.8308 (0.7212,0.8436)	0.8342 (0.7247,0.8445)	0.8420 (0.6381,0.7324)	0.5102 (0.2629,0.4006)

Table 9b. Informed Coverage rates stratified by geographic category, July 1, 2009 – December 31, 2010

	Rural	Urban Cluster	Urbanized Area	Missing Geography
AK	-	-	0.4805 (1.0000,0.0000)	0.8512 (0.8143,0.7878)
AL	0.7825 (0.7598,0.8053)	0.7612 (0.7385,0.7840)	0.7465 (0.3929,0.7237)	0.9289 (0.8117,0.9062)
AR	0.8425 (0.8113,0.8737)	0.6740 (0.6428,0.7052)	0.7261 (0.5921,0.6948)	0.8951 (0.8551,0.8638)
AZ	-	-	0.2655 (1.0000,0.0000)	0.8964 (0.7961,0.8808)
CA	0.6782 (0.6693,0.6871)	0.6779 (0.6690,0.6869)	0.6651 (0.4943,0.6562)	0.8432 (0.7742,0.8342)
CO	0.4734 (0.4281,0.5187)	0.4924 (0.4471,0.5377)	0.4796 (0.4493,0.4342)	0.7347 (0.7313,0.6894)
CT	0.6699 (0.6408,0.6990)	0.6282 (0.5991,0.6573)	0.6253 (0.4860,0.5962)	0.8993 (0.8671,0.8702)
DE	-	-	-	0.8333 (0.8861,0.6612)
FL	0.4293 (0.4095,0.4491)	0.4129 (0.3931,0.4328)	0.4028 (0.3432,0.3830)	0.7257 (0.6996,0.7059)
GA	0.7326 (0.6981,1.0000)	-	0.6275 (0.7820,0.5530)	0.7638 (0.7079,0.7292)
HI	-	-	-	0.9000 (0.9228,0.8070)
IA	0.7967 (0.7458,0.8476)	0.7719 (0.7210,0.8228)	0.6086 (0.4879,0.5577)	0.8455 (0.8057,0.7946)
ID	0.8315 (0.7825,0.8805)	0.8357 (0.7867,0.8846)	0.8355 (0.8054,0.7865)	0.9243 (0.9242,0.8754)
IL	0.9458 (0.9358,0.9559)	0.9477 (0.9376,0.9577)	0.9610 (0.9086,0.9509)	0.8513 (0.8321,0.8413)
IN	0.5734 (0.5395,0.6072)	0.5581 (0.5242,0.5919)	0.5312 (0.3880,0.4973)	0.8758 (0.8513,0.8419)
KS	0.4924 (0.3863,0.5985)	0.5065 (0.4004,0.6126)	0.4751 (0.4630,0.3690)	0.7502 (0.7674,0.6441)
LA	0.9396 (0.9223,0.9569)	0.9432 (0.9259,0.9605)	0.9413 (0.9125,0.9240)	0.9338 (0.9046,0.9165)
MD	0.6738 (0.6439,1.0000)	0.8134 (0.7836,0.9254)	0.7217 (0.8060,0.6806)	0.8694 (0.8398,0.8396)
MI	0.5524 (0.5245,0.5804)	0.5814 (0.5534,0.6093)	0.5427 (0.4930,0.5148)	0.8510 (0.8445,0.8230)
MN	0.4172 (0.3745,0.4599)	0.4390 (0.3963,0.4816)	0.4367 (0.3080,0.3940)	0.8065 (0.7499,0.7638)
MO	0.7429 (0.7162,0.7696)	0.6566 (0.6299,0.6833)	0.6602 (0.5763,0.6335)	0.8869 (0.8399,0.8602)
MT	0.7350 (0.6455,0.8305)	0.7352 (0.6455,0.8306)	0.7367 (0.7834,0.6469)	0.7992 (0.8566,0.7075)
NC	0.8104 (0.7844,0.8364)	0.8123 (0.7863,0.8383)	0.7970 (0.7838,0.7710)	0.7964 (0.7806,0.7704)

	Rural	Urban Cluster	Urbanized Area	Missing Geography
ND	0.9176 (0.8319,0.9218)	0.9043 (0.8186,0.9102)	0.9104 (0.8502,0.8247)	0.8976 (0.8388,0.8118)
NE	0.7562 (0.6975,1.0000)	0.6486 (0.3705,0.9845)	0.6624 (0.7456,0.6037)	0.8400 (0.8362,0.7814)
NH	0.9164 (0.8465,0.9308)	0.9229 (0.8530,0.9374)	0.9157 (0.8843,0.8457)	0.9036 (0.9075,0.8337)
NJ	0.8613 (0.8388,0.8839)	0.8534 (0.8308,0.8760)	0.8746 (0.8339,0.8520)	0.8703 (0.8480,0.8478)
NM	0.9301 (0.9068,1.0000)	0.8487 (0.8254,0.9515)	0.8809 (0.9342,0.8576)	0.9065 (0.8677,0.8832)
NV	-	-	0.3348 (0.6402,0.2476)	0.4167 (0.4891,0.3442)
NY	0.8822 (0.8699,0.8945)	0.8884 (0.8761,0.9007)	0.9015 (0.8256,0.8892)	0.9149 (0.8682,0.9026)
OK	0.5337 (0.5078,0.5595)	0.5882 (0.5624,0.6141)	0.6005 (0.5118,0.5747)	0.8586 (0.8040,0.8327)
OR	0.7767 (0.7404,0.8129)	0.7815 (0.7452,0.8178)	0.7818 (0.7853,0.7455)	0.8359 (0.8487,0.7997)
RI	0.8525 (0.6067,0.8944)	0.6154 (0.3770,0.7133)	0.7635 (0.6026,0.6710)	0.9183 (0.8769,0.8763)
SC	0.4255 (0.3685,0.4871)	0.4047 (0.3551,0.4663)	0.3779 (0.4105,0.3453)	0.7348 (0.7805,0.6891)
SD	0.7563 (0.6838,0.8288)	0.8084 (0.7359,0.8809)	0.7928 (0.5639,0.7203)	0.9052 (0.8523,0.8327)
TN	0.5697 (0.5404,0.5990)	0.5703 (0.5411,0.5996)	0.5786 (0.4443,0.5493)	0.8815 (0.8496,0.8522)
TX	0.5574 (0.5447,0.5700)	0.5685 (0.5558,0.5812)	0.5375 (0.4716,0.5248)	0.7257 (0.6763,0.7130)
UT	0.6408 (0.5658,0.7158)	0.6408 (0.5658,0.7158)	0.6521 (0.6487,0.5771)	0.6661 (0.6707,0.5911)
VA	0.5662 (0.5286,0.7677)	0.4344 (0.2631,0.6318)	0.4135 (0.5845,0.3166)	0.8290 (0.8189,0.7915)
VT	0.9502 (0.9111,0.9526)	0.9534 (0.9181,0.9560)	0.9530 (0.9248,0.9140)	0.9319 (0.9271,0.9154)
WA	0.5988 (0.5794,0.6182)	0.6053 (0.5859,0.6247)	0.6081 (0.4095,0.5887)	0.9048 (0.8682,0.8854)
WI	0.6433 (0.6104,0.6787)	0.5371 (0.5042,0.5699)	0.5982 (0.5948,0.5653)	0.8712 (0.8669,0.8383)
WY	0.8340 (0.7382,0.9100)	0.8223 (0.7265,0.8946)	0.8353 (0.7752,0.7396)	0.8520 (0.8002,0.7563)

To determine which states had adequate data to be used in the Informed Coverage measure the following method was used: Each child with appendicitis was matched to 10 children without appendicitis via Mahalanobis distance optimal matching (Rosenbaum, 2010) with a distance matrix that included age and exact matched on gender, the two most clinically relevant risk factors for appendicitis (Addiss, 1990). This generated a control pool of children that had the same gender and very similar, if not identical, age to their matched counterpart. For each child with appendicitis, to avoid bias of retroactive coverage, a point-in-time four months before the date of appendicitis admission was used to determine whether the child was covered via FFS/PCCM or managed care, and the same month was used for their non-appendicitis matched counterpart. In the context of noninferiority testing (Wellek, 2010), a state was deemed to have insufficient managed care appendicitis claims if the 95% confidence interval for the managed care rate in the appendicitis children

minus the managed care rate in the matched controls was completely below -2%. The six previously mentioned states failed to meet this criterion; the results of this process are seen in the table below.

Table 10: Results of matched analysis comparing reported appendicitis claims for managed care and fee-for-service patients in each state. States were eliminated if the entire 95% CI for the rate of reported claims for appendicitis in managed care plans minus the rate in the matched controls was below -2%. States highlighted in Gray did not meet criteria to be utilized for studying appendectomy claims. Data assessed on calendar year 2008.

State	State Apx N	State Apx rate	Case FFS	Case MC	Control FFS	Control MC	FFS % diff	MC % diff	95% CI
AK	76	0.0945%	63	0	606	0	0.00%	0.00%	NA*
AL	258	0.0559%	180	44	1783	386	-1.85%	1.85%	(-3.33,7.03)
AR	253	0.0595%	212	13	1973	200	3.43%	-3.43%	(-6.73,-0.12)
AZ	943	0.1336%	95	596	988	6046	-0.30%	0.30%	(-2.35,2.94)
CA	3763	0.0853%	333	2658	2925	27436	1.50%	-1.50%	(-2.67,-0.33)
CO	194	0.0652%	23	95	190	1172	5.54%	-5.54%	(-12.95,1.87)
CT	43	0.0167%	21	1	332	47	7.86%	-7.86%	(-20.21,4.50)
DC^	2	0.0027%	0	0	2	15	NA	NA	NA
DE	9	0.0106%	2	6	10	59	10.51%	-10.51%	(-38.21,17.20)
FL	785	0.0469%	107	359	1273	4659	1.50%	-1.50%	(-5.33,2.33)
GA	300	0.0308%	10	202	148	2090	-1.90%	1.90%	(-0.96,4.76)
HI	27	0.0244%	1	23	16	222	-2.56%	2.56%	(-6.55,11.66)
IA	97	0.0392%	53	31	426	349	8.13%	-8.13%	(-19.26,3.00)
ID	140	0.1057%	85	9	964	74	-2.45%	2.45%	(-3.26,8.15)
IL	945	0.0716%	812	67	7432	942	3.63%	-3.63%	(-5.51,-1.74)
IN	218	0.0353%	43	131	357	1388	4.25%	-4.25%	(-10.92,2.41)
KS	129	0.0574%	15	82	177	828	-2.15%	2.15%	(-5.11,9.40)
KY	308	0.0704%	182	87	1519	1054	8.62%	-8.62%	(-14.63,-2.61)
LA	381	0.0547%	350	0	3527	0	0.00%	0.00%	NA
MA	99	0.0236%	63	17	317	464	38.16%	-38.16%	(-47.12,-29.20)
MD	246	0.0518%	15	191	120	2013	1.66%	-1.66%	(-5.38,2.07)
ME^									
MI	223	0.0220%	9	140	39	1880	4.01%	-4.01%	(-7.81,-0.21)
MN	245	0.0616%	52	121	557	1455	2.37%	-2.37%	(-9.51,4.76)
MO	349	0.0645%	122	167	1061	1797	5.09%	-5.09%	(-11.14,0.96)
MS	165	0.0426%	52	83	418	927	7.44%	-7.44%	(-10.31,-4.57)
MT	49	0.0872%	30	0	365	0	0.00%	0.00%	NA
NC	548	0.0661%	387	2	4237	19	-0.07%	0.07%	(-0.67,0.81)
ND	28	0.0727%	21	0	210	0	0.00%	0.00%	NA
NE	79	0.0490%	14	47	192	448	-7.05%	7.05%	(-4.15,18.25)
NH	41	0.0546%	36	0	331	0	0.00%	0.00%	NA
NJ	477	0.0846%	30	364	414	3597	-2.71%	2.71%	(-0.11,5.52)
NM	319	0.1029%	67	204	581	2212	3.92%	-3.92%	(-9.28,1.44)
NV	72	0.0394%	3	25	50	468	1.06%	-1.06%	(-12.81,10.69)
NY	1285	0.0707%	165	984	2278	8826	-6.15%	6.15%	(3.99,8.32)
OH	73	0.0064%	27	10	125	508	53.23%	-53.23%	(-68.14,-38.31)
OK	421	0.0869%	38	317	352	3213	0.83%	-0.83%	(-4.15,2.49)
OR	196	0.0805%	17	125	158	1209	0.41%	-0.41%	(-6.08,5.26)
PA	138	0.0137%	93	9	279	887	67.25%	-67.25%	(-73.63,-60.87)

State	State Apx N	State Apx rate	Case FFS	Case MC	Control FFS	Control MC	FFS % diff	MC % diff	95% CI
RI	83	0.0908%	0	66	100	575	-14.81%	14.81%	NA
SC	160	0.0337%	13	110	163	1150	-1.85%	1.85%	(-3.98,7.67)
SD	45	0.0571%	36	0	370	0	0.00%	0.00%	NA
TN	244	0.0354%	0	214	0	2169	0.00%	0.00%	NA
TX	2929	0.1010%	1074	962	11093	11686	4.05%	-4.05%	(-6.29,-1.81)
UT	95	0.0562%	0	53	14	615	-2.23%	2.23%	NA
VA	204	0.0427%	61	99	500	1150	7.82%	-7.82%	(-15.62,-0.02)
VT	43	0.0802%	39	0	364	0	0.00%	0.00%	NA
WA	443	0.0693%	5	377	11	3717	1.01%	-1.01%	(-2.17,0.14)
WI	238	0.0567%	63	131	564	1202	0.54%	-0.54%	(-7.62,6.54)
WV	32	0.0177%	21	3	66	190	61.72%	-61.72%	(-75.09,-48.34)
WY	31	0.0566%	27	0	243	0	0.00%	0.00%	NA

*States marked "NA" do not have any managed care health plans in place. All claims data in these states is reported via FFS or PCCM, wherein claims are billed as FFS.

^In 2008, the state of Maine was excluded because, due to a lack of a functional MMIS system, they do not report any inpatient claims. The District of Columbia also did not submit a complete dataset to CMS.

- **FFS:** Fee-for-Service, or plans that report claims as FFS, such as Primary Care Case Management
- **MC:** Managed Care, capitated comprehensive managed care plans
- **Matched Case-Control Groups:** Children who had an appendectomy over the course of a calendar year (**cases**) were matched to ten children who did not have an appendectomy (**controls**) on age and gender, and the proportions of MC and FFS in each group were compared in order to ensure the managed care proportion among the appendectomy children is comparable to the proportion in the state as a whole. This provides a check on the completeness of managed care data in each state.

Table 11. Comparison of 95% CI from Boostrap Samples (1,000 per state) and Calculated using Formula using the Validation Time Period

	Coverage PE		Informed PE		Coverage PI		Appendectomy Coverage		Informed Coverage	
	Form.	BStr.	Form.	BStr.	Form.	BStr.	Form.	BStr.	Form.	BStr.
AK	0.8403, 0.8447	0.8410, 0.8439	0.7315, 0.8142	0.7293, 0.8106	0.9193, 0.9228	0.9200, 0.9221	0.7878, 0.9146	0.7872, 0.9128	0.7878, 0.9146	0.7872, 0.9128
AL	0.8226, 0.8246	0.8226, 0.8245	0.7859, 0.8110	0.7852, 0.8103	0.9608, 0.9619	0.9610, 0.9618	0.9060, 0.9515	0.9066, 0.9505	0.9060, 0.9515	0.9066, 0.9505
AR	0.8811, 0.8829	0.8814, 0.8825	0.8142, 0.8547	0.8123, 0.8524	0.9354, 0.9368	0.9357, 0.9365	0.8637, 0.9261	0.8604, 0.9227	0.8637, 0.9261	0.8604, 0.9227
AZ	0.8267, 0.8283	0.8270, 0.8280	0.7777, 0.7961	0.7773, 0.7959	0.9067, 0.9080	0.9069, 0.9077	0.8808, 0.9120	0.8806, 0.9123	0.8808, 0.9080	0.8806, 0.9074
CA	0.8352, 0.8358	0.8353, 0.8357	0.7627, 0.7739	0.7628, 0.7737	0.9107, 0.9112	0.9108, 0.9111	0.8341, 0.8519	0.8345, 0.8519	0.8341, 0.8519	0.8345, 0.8519
CO	0.8476, 0.8495	0.8479, 0.8491	0.6641, 0.7307	0.6639, 0.7291	0.9090, 0.9106	0.9093, 0.9103	0.6889, 0.7796	0.6891, 0.7726	0.6889, 0.7796	0.6895, 0.7726
CT	0.9119, 0.9138	0.9122, 0.9135	0.8188, 0.8658	0.8162, 0.8659	0.9550, 0.9564	0.9552, 0.9562	0.8695, 0.9276	0.8667, 0.9280	0.8695, 0.9276	0.8667, 0.9280
DE	0.8549, 0.8589	0.8554, 0.8584	0.6373, 0.8861	0.6092, 0.8576	0.9232, 0.9264	0.9237, 0.9259	0.6612, 1.0000	0.6471, 1.0000	0.6612, 0.9264	0.6667, 0.9254
FL	0.8531, 0.8540	0.8532, 0.8538	0.6682, 0.6985	0.6683, 0.6978	0.9165, 0.9173	0.9167, 0.9171	0.7049, 0.7446	0.7061, 0.7433	0.7049, 0.7446	0.7061, 0.7433
GA	0.8383, 0.8394	0.8385, 0.8392	0.6545, 0.7079	0.6549, 0.7079	0.9081, 0.9090	0.9083, 0.9089	0.7292, 0.7984	0.7294, 0.7988	0.7292, 0.7984	0.7294, 0.7988
HI	0.9281, 0.9306	0.9285, 0.9302	0.7500, 0.9228	0.7381, 0.9114	0.9626, 0.9646	0.9630, 0.9642	0.8070, 0.9930	0.7941, 0.9808	0.8070, 0.9646	0.7941, 0.9638
IA	0.8719, 0.8742	0.8723, 0.8738	0.7246, 0.8055	0.7243, 0.8055	0.9266, 0.9285	0.9269, 0.9281	0.7945, 0.8962	0.7939, 0.8947	0.7945, 0.8962	0.7939, 0.8947
ID	0.8545, 0.8575	0.8549, 0.8570	0.7474, 0.8127	0.7464, 0.8125	0.9208, 0.9232	0.9213, 0.9227	0.7921, 0.8901	0.7905, 0.8914	0.7921, 0.8901	0.7905, 0.8914
IL	0.9284, 0.9291	0.9285, 0.9290	0.8848, 0.9019	0.8845, 0.9013	0.9650, 0.9656	0.9651, 0.9655	0.9435, 0.9636	0.9435, 0.9631	0.9435, 0.9636	0.9435, 0.9631
IN	0.8865, 0.8878	0.8867, 0.8876	0.8038, 0.8496	0.8041, 0.8479	0.9362, 0.9372	0.9364, 0.9370	0.8408, 0.9085	0.8417, 0.9050	0.8408, 0.9085	0.8417, 0.9050
KS	0.8211, 0.8238	0.8215, 0.8234	0.6208, 0.7671	0.6198, 0.7646	0.9117, 0.9139	0.9121, 0.9134	0.6439, 0.8561	0.6400, 0.8547	0.6439, 0.8561	0.6466, 0.8547
LA	0.9433, 0.9443	0.9435, 0.9441	0.8824, 0.9124	0.8821, 0.9114	0.9783, 0.9789	0.9784, 0.9788	0.9237, 0.9583	0.9226, 0.9576	0.9237, 0.9583	0.9226, 0.9576
MD	0.9003, 0.9018	0.9006, 0.9016	0.7932, 0.8398	0.7922, 0.8387	0.9470, 0.9482	0.9472, 0.9479	0.8396, 0.8992	0.8369, 0.8993	0.8396, 0.8992	0.8369, 0.8993
MI	0.8979, 0.8989	0.8981, 0.8988	0.8055, 0.8442	0.8040, 0.8441	0.9454, 0.9462	0.9455, 0.9460	0.8227, 0.8787	0.8229, 0.8784	0.8227, 0.8787	0.8230, 0.8784
MN	0.7823, 0.7847	0.7823, 0.7846	0.7022, 0.7467	0.7023, 0.7454	0.9503, 0.9517	0.9505, 0.9515	0.7615, 0.8469	0.7604, 0.8443	0.7615, 0.8469	0.7604, 0.8443
MO	0.8764, 0.8779	0.8766, 0.8776	0.8026, 0.8388	0.8018, 0.8383	0.9306, 0.9318	0.9308, 0.9315	0.8593, 0.9127	0.8594, 0.9146	0.8593, 0.9127	0.8594, 0.9146
MT	0.8550, 0.8591	0.8556, 0.8585	0.6510, 0.7881	0.6515, 0.7876	0.9231, 0.9263	0.9237, 0.9257	0.6453, 0.8362	0.6450, 0.8368	0.6510, 0.8362	0.6687, 0.8368
NC	0.8927, 0.8938	0.8929, 0.8936	0.7490, 0.7894	0.7484, 0.7895	0.9385, 0.9394	0.9387, 0.9392	0.7762, 0.8282	0.7755, 0.8307	0.7762, 0.8282	0.7755, 0.8307
ND	0.8453, 0.8511	0.8463, 0.8501	0.7084, 0.8540	0.6961, 0.8480	0.9083, 0.9131	0.9092, 0.9122	0.8353, 1.0000	0.8205, 1.0000	0.8353, 0.9131	0.8205, 0.9120
NE	0.8711, 0.8739	0.8715, 0.8735	0.7576, 0.8362	0.7559, 0.8371	0.9344, 0.9366	0.9348, 0.9362	0.7813, 0.8987	0.7816, 0.8993	0.7813, 0.8987	0.7836, 0.8993
NH	0.8817, 0.8853	0.8823, 0.8847	0.7906, 0.8881	0.7828, 0.8837	0.9261, 0.9292	0.9267, 0.9286	0.8467, 0.9866	0.8333, 0.9808	0.8467, 0.9292	0.8346, 0.9284

NJ	0.9015, 0.9027	0.9017, 0.9025	0.7984, 0.8344	0.7980, 0.8332	0.9522, 0.9531	0.9523, 0.9529	0.8512, 0.8964	0.8490, 0.8965	0.8512, 0.8964	0.8490, 0.8965
NM	0.9066, 0.9084	0.9069, 0.9081	0.8322, 0.8677	0.8306, 0.8671	0.9488, 0.9502	0.9490, 0.9499	0.8832, 0.9298	0.8819, 0.9286	0.8832, 0.9298	0.8819, 0.9286
NV	0.8117, 0.8146	0.8121, 0.8142	0.3442, 0.4891	0.3441, 0.4865	0.8979, 0.9003	0.8983, 0.8999	0.3095, 0.4839	0.3133, 0.4866	0.3442, 0.4891	0.3474, 0.4934
NY	0.8759, 0.8768	0.8760, 0.8766	0.8081, 0.8262	0.8075, 0.8251	0.9346, 0.9353	0.9347, 0.9351	0.8884, 0.9130	0.8879, 0.9128	0.8884, 0.9130	0.8879, 0.9128
OK	0.8576, 0.8593	0.8579, 0.8590	0.7682, 0.8037	0.7671, 0.8037	0.9169, 0.9183	0.9172, 0.9181	0.8325, 0.8842	0.8326, 0.8853	0.8325, 0.8842	0.8326, 0.8853
OR	0.8518, 0.8538	0.8521, 0.8534	0.7436, 0.7891	0.7421, 0.7877	0.9147, 0.9164	0.9150, 0.9160	0.7487, 0.8213	0.7508, 0.8223	0.7487, 0.8213	0.7540, 0.8223
RI	0.8677, 0.8714	0.8683, 0.8708	0.8480, 0.8766	0.8432, 0.8705	0.9166, 0.9197	0.9173, 0.9191	0.8997, 0.9836	0.8943, 0.9822	0.8997, 0.9197	0.8943, 0.9190
SC	0.8867, 0.8882	0.8870, 0.8880	0.6882, 0.7795	0.6877, 0.7760	0.9357, 0.9370	0.9360, 0.9367	0.6500, 0.7731	0.6507, 0.7737	0.6882, 0.7795	0.6880, 0.7808
SD	0.8596, 0.8637	0.8601, 0.8630	0.7357, 0.8508	0.7273, 0.8461	0.9221, 0.9254	0.9226, 0.9248	0.8323, 0.9772	0.8254, 0.9701	0.8323, 0.9254	0.8254, 0.9244
TN	0.9038, 0.9051	0.9041, 0.9048	0.8025, 0.8486	0.8038, 0.8467	0.9503, 0.9513	0.9505, 0.9511	0.8516, 0.9101	0.8498, 0.9078	0.8516, 0.9101	0.8498, 0.9078
TX	0.8068, 0.8076	0.8069, 0.8074	0.6589, 0.6762	0.6582, 0.6757	0.8849, 0.8856	0.8850, 0.8855	0.7130, 0.7383	0.7132, 0.7371	0.7130, 0.7383	0.7132, 0.7371
UT	0.7929, 0.7957	0.7934, 0.7952	0.5402, 0.6487	0.5397, 0.6475	0.8760, 0.8784	0.8764, 0.8780	0.5766, 0.7266	0.5776, 0.7267	0.5766, 0.7266	0.5777, 0.7267
VA	0.9029, 0.9042	0.9031, 0.9040	0.7588, 0.8189	0.7576, 0.8184	0.9464, 0.9474	0.9466, 0.9472	0.7915, 0.8666	0.7872, 0.8671	0.7915, 0.8666	0.7872, 0.8671
VT	0.9159, 0.9197	0.9165, 0.9190	0.9159, 0.9197	0.9165, 0.9190	0.9492, 0.9524	0.9498, 0.9517	0.9238, 1.0000	0.9186, 1.0000	0.9238, 0.9524	0.9189, 0.9516
WA	0.9062, 0.9074	0.9064, 0.9072	0.8383, 0.8667	0.8380, 0.8662	0.9533, 0.9542	0.9534, 0.9540	0.8845, 0.9234	0.8852, 0.9240	0.8845, 0.9234	0.8852, 0.9240
WI	0.9001, 0.9016	0.9004, 0.9014	0.8234, 0.8660	0.8226, 0.8650	0.9467, 0.9478	0.9469, 0.9476	0.8374, 0.9032	0.8364, 0.9011	0.8374, 0.9032	0.8385, 0.9011
WY	0.8188, 0.8244	0.8198, 0.8236	0.6209, 0.7717	0.6171, 0.7619	0.8972, 0.9020	0.8982, 0.9012	0.7348, 0.9262	0.7327, 0.9107	0.7348, 0.9020	0.7327, 0.8998

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3166
Measure Title: Antibiotic Prophylaxis Among Children with Sickle Cell Anemia
Measure Steward: QMETRIC - University of Michigan
Brief Description of Measure: The percentage of children ages 3 months to 5 years old with sickle cell anemia (SCA, hemoglobin [Hb] SS) who were dispensed appropriate antibiotic prophylaxis for at least 300 days within the measurement year.
Developer Rationale: Children with SCA are at an increased risk of infection compared with children without the disorder. Daily receipt of antibiotic prophylaxis substantially reduces the risk of infection among these high-risk children. Prior studies indicate that appropriate antibiotic prophylaxis rates are low; however, these reports are limited in their generalizability, as they are usually focused on a single healthcare provider or registry. This measure establishes a claims-based method for identifying appropriate antibiotic prophylaxis among larger and broader populations of children with SCA. The performance scores calculated through this measure will identify areas in need of improvement of antibiotic prophylaxis among children with SCA.

Numerator Statement: The numerator is the number of children ages 3 months to 5 years old with SCA (Hb SS) who were dispensed appropriate antibiotic prophylaxis for at least 300 days within the measurement year.
Denominator Statement: The denominator is the number of children ages 3 months to 5 years with SCA (Hb SS) within the measurement year.
Denominator Exclusions: There are no denominator exclusions.

Measure Type: Process
Data Source: Claims (Only)
Level of Analysis: Health Plan

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

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

The developer provides the following evidence for this measure:

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

Evidence Summary

- The measure has the following logic model: Daily receipt of appropriate antibiotic prophylaxis results in substantial reduction of the incidence of infection among children with sickle cell anemia (SCA).

- The measure developer provided two key sources of evidence. The first was a systematic evidence review and clinical practice guidelines published by the National Heart, Lung, and Blood Institute: Evidence-Based Management of Sickle Cell Disease in 2014. The panel recommended that all children with HbSS be administered oral penicillin prophylaxis (125 mg for age <3 years and 250 mg for age ≥3 years) twice daily until age (Strong Recommendation, Moderate-Quality Evidence) based on evidence from three RCTs and one observational study. The grade assigned was strong recommendation, moderate quality evidence.
- In addition, a Cochrane Systematic Review, published in 2014, found that prophylactic penicillin significantly reduces risk of pneumococcal infection in children with homozygous sickle cell disease, and is associated with minimal adverse reactions. The review included three trials including data from over 800 children. Evidence included in the review was not graded.

Questions for the Committee:

- *What is the relationship of this measure to patient outcomes?*
- *How strong is the evidence for this relationship?*
- *Is the evidence directly applicable to the process of care being measured?*
- *What is the importance of the timeframe the evidence was generated (studies included in the systematic reviews span from 1984- 1995)*?

Guidance from the Evidence Algorithm: Process measure based on systematic review (Box 3)→ QQC presented (Box 4)→ Quantity: moderate; Quality: moderate; Consistency: high (Box 5)→ Moderate (Box 5b)→ Moderate

The highest possible rating is HIGH.

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

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

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

- Measure scores for the measure as specified across six states from 2005-2010 supported findings from prior studies indicating that appropriate antibiotic prophylaxis rates are low—ranging from 5.71% (South Carolina, 2009) to 36.11% (Illinois, 2007). Sample sizes are small, ranging from populations of 35-282 children with sickle cell per state per year.
- 2010 performance scores in the 6 states ranged from 15.6% (Florida) to 27.9% (Texas).

Disparities

- The dataset used for performance scores is Medicaid; therefore, there are no disparities identified by insurance or socioeconomic status. The majority of children with sickle cell disease (approximately 80%) have been enrolled in Medicaid at some point in time. A study assessing compliance with penicillin prophylaxis for sickle cell disease found that adherence was significantly greater in patients with private versus public insurance (17/28 [61%] vs. 33/90 [37%], respectively) (Teach et al., 1998.)—however, disparities in between type of insurance would not be captured by this measure.

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*
- *Are you aware of evidence of other disparities exist in this area of healthcare?*

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

Committee pre-evaluation comments

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

- This is a process measure. This measure (i.e. antibiotic prophylaxis) will result in better health outcomes (e.g. less infections.) Although this measure will examine daily administration, it is unclear how it will be determined if additional (i.e. not prophylactic) antibiotics are administered for breakthrough infections. In addition, the developer used two sources of evidence related to penicillin prophylaxis yet other antibiotics such as erythromycin, sulfamethoxazole, or trimethoprim could be used. The systemic review looks thorough and recent (2014). Rating: moderate
- The evidence provided is based on the Clinical practice guideline from National Heart, Lung and Blood Institute and a Cochrane Systematic Review. Based on this evidence, improvement in this process should lead to improvement in patient outcomes (decrease in pneumococcal infection). The Clinical practice guideline was graded as a strong recommendation and is directly applicable to the process being measured.
- Although it is noted that prophylaxis rates were low, there was also a concern noted due to small sample size. In addition, although there are not disparities in insurance as sampling was from Medicaid, there was nothing on racial/ethnic disparities for African American, Hispanic, and Asian children. Rating: moderate.
- There appears to be a significant gap in care with reported prophylaxis levels only between 5 and 36%, depending on the state. There is significant room to improve performance in this measure. As all patients in performance data were Medicaid patients, there were no further subgroup data provided.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

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

Data source(s): Claims

Specifications:

- Level of analysis: Health Plan
- Care setting: Settings represented with prescription medication claims data
- Interpretation of Score: Better quality = higher score
- Numerator: *number of children ages 3 months to 5 years old with SCA (Hb SS) who were dispensed appropriate antibiotic prophylaxis for at least 300 days within the measurement year.*
- Denominator: *number of children ages 3 months to 5 years with SCA (Hb SS) within the measurement year.*
  - Children are identified through the presence of at least three separate healthcare encounters related to Hb SS within the measurement year identified through either ICD-9 or ICD-10 codes.
- Exclusions: There are no denominator exclusions; however, children with SCA are included starting at 3 months of age to account for any lag in identification and confirmation of the sickle cell disease status of the child.
- The developer includes a calculation algorithm.
- This measure is not risk-adjusted.
- This measure does not involve sampling as all cases are used, and no data abstraction tool is provided.

Questions for the Committee:

- *Is the identification of the target population though the presence of at least three separate healthcare encounters related to Hb SS within the measurement year appropriate?*
- *Are there any children who should be excluded?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing, Testing attachment

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

SUMMARY OF TESTING

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

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

Method(s) of reliability testing:

- Analysis was conducted using Medicaid claims reported to CMS for Medicaid enrollees within the state of Michigan (2007-2011) as well as MAX data for all Medicaid claims reported to CMS for Medicaid enrollees within six state Medicaid programs with moderate to high prevalence of sickle cell anemia: Florida, Illinois, Louisiana, Michigan, South Carolina and Texas (2005-2010).
- The reliability of the measure was tested in the MAX data using a signal-to-noise analysis focused on assessing the reliability to distinguish the performance of one state's Medicaid program from that of another state. Reliability was estimated with a beta-binomial model.

Results of reliability testing: State-specific reliability results for appropriate antibiotic prophylaxis among children with sickle cell anemia are detailed in Table 3. These results show that the reliability based on signal-to-noise analysis ranged from 0.83 to 0.96, with a median of 0.89.

Table 3. State-specific reliability for measure

State	Numerator	Denominator	Reliability
Florida	163	1145	0.9592
Illinois	125	447	0.8475
Louisiana	142	687	0.9130
Michigan	67	432	0.8919
South Carolina	48	273	0.8251
Texas	173	642	0.8909
Median (range)			89.14% (82.51-95.92)

- State-specific reliability was high; observed reliability was consistently greater than 0.80. 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 (RAND Corporation, TR-653-NCQA, 2009).
- The median reliability observed across states was 0.89 (range: 0.83-0.96), which is consistent with a high degree of reliability.

Questions for the Committee:

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

Guidance from the Reliability Algorithm: Empirical reliability testing conducted (Box 2)→ Computed performance measure scores presented (Box 4)→ Appropriate method described: signal-to-noise (Box 5)→ High certainty that the performance measure scores are reliable (Box 6a)→ High

The highest possible rating is HIGH.

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

2b. Validity

2b1. Validity: Specifications

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

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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

SUMMARY OF TESTING

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

Method of validity testing of the measure score:

- ☒ **Face validity**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- **Data element testing against a gold standard:**
  - **Numerator:** The accuracy of administrative claims in identifying antibiotic prophylaxis was assessed through comparison with the gold standard of medical charts. An audit was conducted by medical record abstractors to compare administrative claims data with medical records data. 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.
  - **Denominator**: 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).
- **Empirical Validity Testing of Performance Measure:** The developer tested empirical validity by comparing the MAX data for the state of Michigan (obtained from CMS) with the gold standard of Michigan Medicaid data (obtained directly from Michigan's claims data warehouse) for the same time period (2007-2010). Rates of antibiotic prophylaxis 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.
- **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 (QMETRIC).

Validity testing results:

- **Data element testing against a gold standard:**
  - **Numerator:** For this comparison, 34 children with sickle cell anemia who were enrolled within Michigan Medicaid were successfully matched with their Michigan Medicaid administrative claims data. Eighteen charts also were chosen for calculation of inter-rater reliability; the two trained abstractors had 100% agreement with each other for abstracting antibiotic prescriptions from the medical records. The majority (33/34, 97%) of prescribed antibiotics in the medical record were reflected in administrative claims. Further,

30/34 (88%) of those who were prescribed an antibiotic in the medical record were dispensed their antibiotic within 30 days of the prescription.

- **Denominator:** For this comparison, 865 children met eligibility criteria in 2010 (at least one SCD-related claim ages 1-18, continuous enrollment in Michigan Medicaid in 2010, a NBS result available); 836 children met eligibility criteria in 2011. In 2010, a case definition of three Hb SS 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 the study population in 2011 (see Table 4).

Table 4. Accuracy of case definition of at least 1, 2 and 3 Hb SS claims within a year to identify children with sickle cell anemia compared with the gold standard of newborn screening

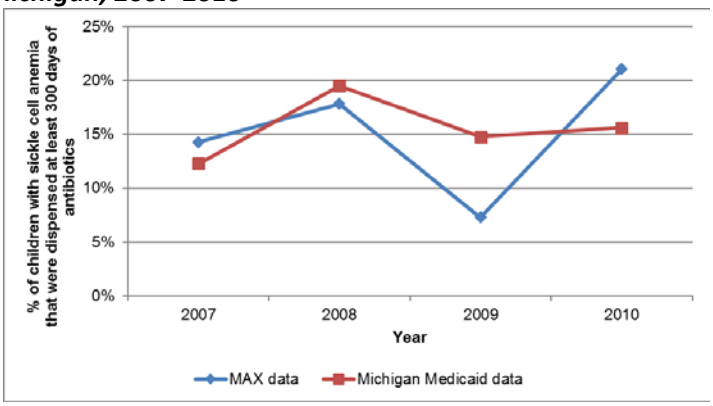
Algorithm	Area under the ROC Curve	# True Positives	# False Positives	# True Negatives	# False Negatives	Sensitivity	Specificity	PPV	NPV
Results - 2010									
≥1 Hb SS Claim	0.50	409	456	0	0	100.0%	0.0%	47.3 %	NA
≥2 Hb SS Claims	0.82	391	144	312	18	95.6%	68.4 %	73.1 %	94.5%
≥3 Hb SS Claims	0.86	374	91	365	35	91.4%	80.0 %	80.4 %	91.3%
Results - 2011									
≥1 Hb SS Claim	0.50	397	439	0	0	100.0%	0.0%	47.5 %	NA
≥2 Hb SS Claims	0.79	377	163	276	20	95.0%	62.9 %	69.8 %	93.2%
≥3 Hb SS Claims	0.87	363	97	342	34	91.4%	77.9 %	78.9 %	91.0%

- **Empirical Validity Testing of Performance Measure:**
  - The comparison of rates of appropriate antibiotic prophylaxis from the gold standard of Michigan Medicaid data compared with MAX data can be seen in Table 5. This illustrates that the number of children who were dispensed at least 300 days of antibiotics among children with sickle cell anemia ranged from 12.28% (2009) to 19.49% (2008) in the claims acquired directly from the Medicaid data warehouse, versus a range of 7.29% (2009) to 21.05% (2010) from MAX data for the same four year period.
  - ***Table 5: Comparison of appropriate antibiotic prophylaxis by source of Medicaid claims data for the state of Michigan, 2007-2010***

Source	Rate Components	2007	2008	2009	2010
Michigan Medicaid data	Numerator	14	23	22	22
	Denominator	114	118	149	141
	Percentage	12.28%	19.49%	14.77%	15.60%
MAX data	Numerator	10	13	7	20
	Denominator	70	73	96	95
	Percentage	14.29%	17.81%	7.29%	21.05%

  - Figure 1 illustrates the performance scores observed between the Michigan Medicaid data from the state warehouse and MAX data from CMS for each overlapping year noted, respectively: 12.3% versus 14.3% (2007); 19.5% versus 17.8% (2008); 14.8% versus 7.3% (2009), and 15.6% versus 21.1% (2010).

Figure 1: Comparison of appropriate antibiotic prophylaxis by source of Medicaid claims data for the state of Michigan, 2007-2010



- Table 6 reports the z-scores and p-values from the two-sample z-tests comparing the proportion of children who were dispensed at least 300 days of antibiotic prophylaxis each year between Michigan Medicaid and MAX data. When antibiotic prophylaxis was assessed for the same state (Michigan) from these two data sources for the same time period (2007-2010), no differences in rates were observed (all p-values >0.05).
- ***Table 6: Comparison of performance score by source of Medicaid claims data, Michigan***

	2007	2008	2009	2010
z-score	0.3921	-0.2890	-1.7678	1.0735
p-value	0.6965	0.7718	0.0767	0.2846

- **Face Validity of Performance Measure Score:** The QMETRIC 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.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:
N/A: No exclusions

Questions for the Committee:

- *Is the lack of exclusions consistent with the evidence?*

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

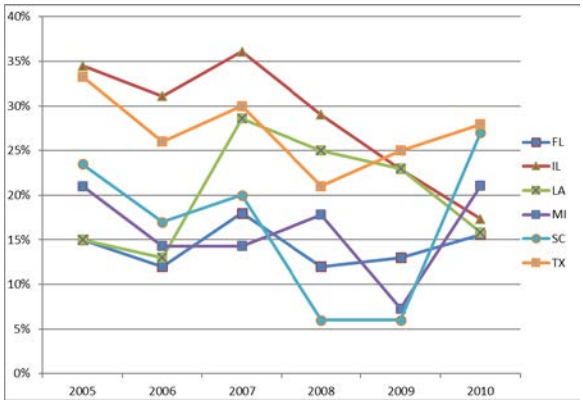
Questions for the Committee:

- *Do you agree with the decision not to risk adjust this process measure?*

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

The proportion of children who were dispensed at least 300 days of antibiotics ranged from 3.23% to 36.11% (Figure 2).

Figure 2. Provision of antibiotic prescriptions, 300-day rate, for children with sickle cell anemia using MAX data, 2005-2010



Compared with 2005, children did not have higher odds of being dispensed at least 300 days of antibiotic prophylaxis (Table 7). However, rates did vary by state (Table 8).

Table 7. Odds of being dispensed at least 300 days of antibiotic prophylaxis among children with sickle cell anemia enrolled in six state Medicaid programs by year using MAX data, 2005-2010

Year	Odds Ratio	95% Confidence Interval
2005	Reference	Reference
2006	0.74	0.48, 1.13
2007	0.94	0.62, 1.41
2008	0.72	0.47, 1.09
2009	0.42	0.26, 0.69
2010	0.83	0.55, 1.28

Table 8. Odds of being dispensed at least 300 days of antibiotic prophylaxis among children with sickle cell anemia enrolled in Medicaid programs by state using MAX data, 2005-2010

State	Odds Ratio	95% Confidence Interval
FL	Reference	Reference
IL	3.12	1.95, 4.97
LA	1.73	1.01, 2.77
MI	1.60	0.94, 2.70
SC	1.33	0.71, 2.50
TX	3.37	2.19, 5.19

Although this measure is successfully able to distinguish differences in performance across states and years, the rates of appropriate antibiotic prophylaxis are not increasing over time.

Question for the Committee:

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

2b6. Comparability of data sources/methods:

- N/A: only one set of specifications

2b7. Missing Data

- The developer tested the completeness and accuracy of Medicaid claims for sickle cell patients seen at three major medical centers in Michigan by matching 34 children with their Medicaid administrative claims data. The developer found that "the vast majority of prescribed antibiotics in the medical record were reflected in

Medicaid claims (33/34, 97%)" and that "only one case (3%) that did not have an antibiotic prescription claim filled within 90 days of a prescription date."
- The developer states that missing data is unlikely to bias the results.

Guidance from the Validity Algorithm: Measure specifications consistent with evidence (Box 2) → Potential threats to validity relevant to the measure assessed (Box 3)→ Empirical validity testing using the measure as specified and the appropriate statistical test (Box 6)→ Validity testing conducted with computed performance measure scores (Box 7)→ Method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships (Box 8)→ Results and scope of testing show there is a high certainty that the performance measure scores are a valid indicator of quality (Box 8a)→ Rate as High.

The highest possible rating is HIGH.

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

Committee pre-evaluation comments

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

- There is no risk adjustment. It is uncertain as to whether these children will have 3 health care encounters, which is a requirement of the target population, as the condition varies in severity. Some exclusions to consider could be those with suppressed immune systems due to comorbid conditions such as organ transplant, cancer, or on other immunosuppressive medications such as steroids, Humira, etc. The calculation algorithm requires continuous enrollment in the health care plan which would not take into consideration children who may go off and on private plans.
- By identifying the target population through 3 separate visits in one year, it seems that those children with sickle cell disease but less than 3 healthcare visits will be excluded. These children may be at higher risk for sepsis.
- It is noted that the minimum reliability level and median reliability were both met. Generalization is questionable due to the small sample size of the empirical evidence. Rating: moderate.
- There were hundreds of patients from each state measured and reliability was consistently > 80% for all states. This size test sample does seem sufficient to generalize for widespread implementation.
- It was reassuring that there was empirical validity testing of the measure score in addition to face validity. The abstractors were consistent in their findings and it is noted that the records used had 97% accuracy. It is noted in this section that the amount considered for prophylaxis would be 300 days. There is some concern as these sample sizes are smaller than those used in the evidence, which were also considered small.
- The fairly tight correlation between MAX data and Medicaid data suggests that the results demonstrate sufficient validity. In addition, the high correlation between the antibiotics prescribed in the chart and administrative claims for those antibiotics suggests the score from this measure is a good indicator of quality.
- This answer is for 2b3-2b7 threats to validity. There is the remaining concern of no risk adjustment or exclusions for comorbid conditions. This is also concern about not distinguishing differences between states and that the data isn't longitudinal. Rating: moderate.
- There was no significant missing data that would threaten the validity of this measure.

Criterion 3. Feasibility

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

The developer provided the following information:

- The data elements required for the measure are routinely generated and used during care delivery and all data elements used in the measure are defined fields in electronic claims.
- The primary information needed for this measure includes basic demographics, diagnosis codes, and procedure codes and dates. These data are widely available, although obtaining them often requires a restricted-use data agreement.
- For multiple-state comparisons, MAX data are available from CMS. When the measure is used at the single-state level, state health departments can use their own Medicaid data.

- While QMETRIC testing efforts support the feasibility of implementing this measure, the testing process demonstrated the technical challenges involved in identifying appropriate SCD cases from very large administrative claims files such as MAX data.
- This measure was also validated using information acquired from medical records from Children's Hospital of Michigan (where each patient had both a paper record and an EHR), and from Hurley Medical Center and the University of Michigan Health System (both institutions had only EHR information). However, EHR systems did not support structured queries of the relevant data; chart abstraction was also necessary. The abstraction process demonstrated that, while the required data for this measure can be obtained from both types of systems, EHRs provided greater efficiency in that they could be reviewed more quickly and remotely if structured queries were possible.

Questions for the Committee:

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

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

Committee pre-evaluation comments

Criteria 3: Feasibility

- n/a
- Patient encounters are easily obtainable using ICD-10 codes through EHR and Medicaid claims data should be easily available electronically. There should be no significant barriers to data collection.

Criterion 4: Usability and Use

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

Current uses of the measure: None
Publicly reported? ☐ **Yes** ☒ **No**

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

Accountability program details:
This is a new measure so it is not currently in use. However, the developer states that "Under our recently awarded PQMP grant, the New York Medicaid program has partnered with QMETRIC to formally begin field testing and implementation of this measure in 2017 at the state, plan, and health system levels." It is unclear, however, whether performance results will be used for accountability (i.e., public reporting or payment).

Improvement results: This is a new measure and improvement information was not provided.

Unexpected findings (positive or negative) during implementation:
The developer reported that no unintended consequences were identified during testing.

Potential harms: No unintended negative consequences to individuals or populations were identified during testing. The Cochrane Review stated "Adverse drug effects were rare and minor. However, there were problems with children keeping to the treatment schedule and with the development of antibiotic resistance."

Vetting of the measure

The measure has not been vetted.

Feedback:
No information provided.

***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 the potential unintended consequences, such as the development of antibiotic resistance?*

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

Committee pre-evaluation comments

Criteria 4: Usability and Use

- n/a
- This measure is not currently publicly reported or being used as a quality measure. This performance measure has the potential to improve high quality care by reducing the burden of preventable pneumococcal infections. The potential consequence of increased antibiotic resistance was mentioned, but not quantified.

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

RATIONALE IF NOT ELIGIBLE:

The measure has not been vetted.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

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

Measure Title: Antibiotic Prophylaxis Among Children with Sickle Cell Anemia

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

Date of Submission: 12/6/2016

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

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

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☐ Health outcome:

☐Patient-reported outcome (PRO):

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

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

☒ Process: Antibiotic Prophylaxis Among Children with Sickle Cell Anemia

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Daily receipt of appropriate antibiotic prophylaxis results in substantial reduction of the incidence of infection among children with sickle cell anemia (SCA).

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

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

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

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

X Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	National Heart, Lung, and Blood Institute. Evidence-Based Management of Sickle Cell Disease: Expert Panel Report, 2014; https://www.nhlbi.nih.gov/health-pro/guidelines/sickle-cell-disease-guidelines.
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	**Recommendations** 1. Administer oral penicillin prophylaxis (125 mg for age <3 years and 250 mg for age ≥3 years) twice daily until age 5 in all children with HbSS. ***(Strong Recommendation, Moderate-Quality Evidence)*** 2. Discontinue prophylactic penicillin in children with HbSS at age 5 unless they have had a splenectomy or invasive pneumococcal infection. When discontinuing penicillin prophylaxis at age 5, it is important to assure that the child has completed the recommended pneumococcal vaccination series, and if not, complete the series immediately. ***(Weak Recommendation, Moderate-Quality Evidence)*** 3. Consider withholding penicillin prophylaxis from children with HbSC disease and HbSβ⁺-thalassemia unless they have had a splenectomy ***(Weak Recommendation, Low-Quality Evidence)*** 4. Assure that people of all ages with SCD have been vaccinated against *Streptococcus pneumoniae*.* ***(Strong Recommendation, Moderate-Quality Evidence)*** 5. Remind people with SCD, their families, and caregivers to seek immediate medical attention whenever fever (temperature greater than 101.3°F or 38.5°C) occurs, due to the risk for severe bacterial infections. ***(Consensus–Panel Expertise)*** * Refer to the "Immunization" section of this chapter for comprehensive information on immunizations.
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	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.

Provide all other grades and definitions from the evidence grading system

Exhibit 4. 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 likely to change the estimate.
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.
Weak recommendation Moderate-quality evidence	Benefits closely balanced with harms and burdens	Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect or imprecise evidence), or unusually strong evidence from unbiased observational studies	Alternative approaches likely to be better for some patients under some circumstances. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Weak recommendation Low-quality evidence	Uncertainty in the estimates of benefits, harms, and burdens; benefits may be closely balanced with harms and burdens	Evidence for at least one critical outcome from observational studies, from RCTs with serious flaws, or indirect evidence	Other alternatives may be equally reasonable. Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Weak recommendation Very low-quality evidence	Major uncertainty in the estimates of benefits, harms, and burdens; benefits may or may not be balanced with harms and burdens	Evidence for at least one critical outcome from unsystematic clinical observations or very indirect evidence	Other alternatives may be equally reasonable. Any estimate of effect, for at least one critical outcome, is very uncertain.

Source: Reprinted with permission of the American Thoracic Society. Copyright © 2012 American Thoracic Society. Schünemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A, Fahy BF, Gould MK, Horan KL, Krishnan JA, Manthous CA, Maurer JR, McNicholas WT, Oxman AD, Rubenfeld G, Turino GM, Guyatt G; ATS Documents Development and Implementation Committee. An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med.* 2006 Sep 1;174(5):605-14. Official Journal of the American Thoracic Society.[29]

* Exceptionally strong evidence from unbiased observational studies includes: (1) evidence from studies that yield estimates of the treatment effect that are large and consistent; (2) evidence in which all potential biases may be working to underestimate an apparent treatment effect, and therefore, the actual treatment effect is likely to be larger than that suggested by the study data; and (3) evidence in which a dose-response gradient exists

Grade assigned to the **recommendation**

Strong Recommendation; for definition, see evidence grading system table above

with definition of the grade	
Provide all other grades and definitions from recommendation grading system	See above; included within evidence grading system table
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Three RCTs and one observational study were included. As stated within the NHLBI Clinical Guidelines: "The three RCTs [Gaston et al., 1986; Faletta et al., 1995; John et al., 1984 {see References at the end of this documents}] were of moderate methodological quality and compared penicillin to no prophylaxis. The initiation of penicillin prophylaxis was associated with a significant reduction in the risk for developing serious pneumococcal infections (2/105 vs. 13/110) and a nonsignificant reduction in mortality (0/105 deaths vs. 3/110 deaths; very low-quality evidence due to severe imprecision). A single trial evaluated the consequences of discontinuing penicillin prophylaxis; it suggested that prophylaxis in children who have not had a prior severe pneumococcal infection or a splenectomy may be discontinued at age 5. Children who continued penicillin had a nonsignificant reduction in systemic pneumococcal infections; there was no effect on mortality. The observational study [Nkouwap et al., 1999] compared penicillin to spiramycin and demonstrated that penicillin was superior. However, the penicillin group had a higher rate of pneumococcal vaccination, confounding the effect of antibiotics and making strong conclusions difficult. The quality of evidence is very low due to severe imprecision (i.e., small number of events) and methodological limitations. Evidence is lacking in children with genotypes other than SS, even though many clinicians prescribe prophylactic penicillin for them both before and after age 5." Please note that this measure specification deviates slightly from the NHLBI recommendations in two main areas: child age and type of antibiotic dispensed. Instead of including children from birth, this measure includes children starting at 3 months of age. This is to account for any lag in identification of the sickle cell disease status of the child at the State level. Although NHLBI guidelines specifically recommend penicillin for antibiotic prophylaxis, some children may have or be suspected to have penicillin sensitivity. The American Academy of Pediatrics Section on Hematology/Oncology and Committee on Genetics suggests an alternative for children who are allergic to penicillin: "Erythromycin prophylaxis may be used as an alternative for children with suspected or proven penicillin allergy." (American Academy of Pediatrics 2002 and 2016). Therefore, we have included a broader definition of appropriate antibiotics in this measure. These modifications to the guidelines are intended to avoid underestimation of the proportion of children with SCA who are protected against pneumococcal infection.

Estimates of benefit and consistency across studies	The four studies detailed within the NHLBI evidence tables (Gaston et al., 1986; Faletta et al., 1995; John et al., 1984; Nkouwap et al., 1999) included a total of 951 children under the age of 5 years. The majority of children (94%) were sickle cell subtype Hb SS (sickle cell anemia); 5% were Hb SC, and 1% were Hb S beta zero thalassemia. All studies investigating the relationship between antibiotic prophylaxis and infection indicated a net benefit to antibiotic prophylaxis among children with sickle cell anemia. The primary study (Gaston et al., 1986) enrolled children with Hb SS starting at 3-6 months of age into the Prophylactic Penicillin Study (PROPS) to test the protective effect of regular, daily administration of oral penicillin against incidence of documented septicemia due to s. pneumonia in children under the age of 3 years. Children were randomized into study groups; participants were allocated to either receive 125mg of penicillin, twice daily (n=105 children), or a placebo (n=110 children), for a mean of 15 months. Reduction of septicemia was found in children who received the penicillin, as an 84% reduction in the incidence of infection was observed in the group treated with penicillin compared with the group given placebo (13 of 110 patients vs. 2 of 105; P = 0.0025).
What harms were identified?	Gaston et al. (1986) reported no adverse effects throughout the course of the study. One child in Falletta et al. (1995) experienced nausea and vomiting on a higher dose of penicillin than explored in the PROPS trial. Given the severity of infection and the potential for increased mortality due to pneumococcal infection among these high risk children, the benefit outweighs the potential harm for this recommendation.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	N/A

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	Hirst C, Owusu-Ofori S. Prophylactic antibiotics for preventing pneumococcal infection in children with sickle cell disease. Cochrane Database of Systematic Reviews 2014, Issue 11. Art. No.: CD003427. DOI: 10.1002/14651858.CD003427.pub3. http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD003427.pub3/abstract;jsessionid=03720C55FB06B08BFB3D12EBCEDF7614.f04t02
Quote the guideline or recommendation verbatim about	"Prophylactic penicillin significantly reduces risk of pneumococcal infection in children with homozygous sickle cell disease, and is associated with minimal adverse reactions."

the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	N/A
Provide all other grades and definitions from the evidence grading system	N/A
Grade assigned to **recommendation** with definition of the grade	N/A
Provide all other grades and definitions from the recommendation grading system	N/A
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	Three trials were included in the review. As stated within the Cochrane Review: "Methodological quality was assessed based on a method described by Schulz et al. (1995). "In the John trial participants were randomised, but no details were given of the method of randomisation (John et al., 1984). The publication reports that the group allocation was changed due to the protocol for injected penicillin prophylaxis groups being inconvenient to some families who lived at remote addresses, or due to age of participants at recruitment so that the duration of penicillin treatment would have been too short to assess. Sixteen participants (6.6%) were therefore reassigned to groups which did not receive penicillin prophylaxis. The groups were uneven, with significantly more participants in the penicillin groups (143 in penicillin group compared to 99 in control group). Full baseline data for participant characteristics were not given. The trial was not blinded. Intention-to-treat analysis was undertaken after participants were reassigned. There were 25

	withdrawals, 20 due to splenectomy, four due to emigration and one because of recurrent meningitis. The participants discontinued penicillin at the age of three years. However, they continued to be analysed in the groups to which they were randomised. Therefore, although there were no cases of pneumococcal infection amongst participants taking penicillin, there were seven cases in the penicillin assigned groups, all occurring after discontinuation of the drug. "In the first PROPS trial, a central coordinating center generated a blocked randomisation sequence, and directed participant entry assignment over the telephone (PROPS, 1986 [Gaston et al., 1986; Gaston and Verter, 1990]). Sealed envelopes were also held at the clinical centres in case the central office could not be reached, to maintain allocation concealment. The participants and centre personnel were blinded to allocation, and placebo tablets looked almost identical to penicillin. A sample size calculation was performed based on an estimated 50% reduction in risk of infection, and as a result 219 participants were recruited from 23 centres throughout the USA. Four participants subsequently withdrew due to revisions of diagnosis of genotype; these patients had no severe infections but were not included in subsequent analyses. The baseline characteristics of the children in each group, including history of palpable spleen or infection, were similar. The trial was terminated early due to extreme results. Because of this, there is a possibility that the reported results may be over-estimated. "In a further PROPS trial, randomisation was by permuted block method, stratified by clinical site and years of previous penicillin use (PROPS II, 1995 [Faletta et al., 1995; Bjornson et al., 1996; Woods et al., 1997]). It was unclear whether allocation concealment had been performed. Identical placebo tablets were used to maintain double blinding of the participants and centre personnel. A sample size calculation ensured that the trial was powered to show a three-fold increase in infection in the placebo group, based on an estimated incidence of 4% in the penicillin group. Four hundred participants were subsequently recruited from 18 centres in the USA. The characteristics of participants in each group were similar at baseline. Four children died after randomisation, but other withdrawals are not reported, and it is unclear whether an intention-to-treat analysis was undertaken."
Estimates of benefit and consistency across studies	"Three trials with over 800 children are included in the review. All three trials showed a reduced rate of infection in children with sickle cell disease receiving penicillin preventatively. Two trials looked at whether treatment was effective. The third trial followed on from one of the early trials and looked at when it was safe to stop treatment. … "All of the included trials showed a reduced incidence of infection in children with sickle cell disease (SS or S beta zero thalassemia) receiving prophylactic penicillin. In trials which investigated initiation of penicillin on risk of pneumococcal infection, the odds ratio was 0.37 (95% CI 0.16 to 0.86), while for withdrawal the odds ratio was 0.49 (95% CI 0.09 to 2.71)."

What harms were identified?	"Adverse drug effects were rare and minor. However, there were problems with children keeping to the treatment schedule and with the development of antibiotic resistance." Note: The American Academy of Pediatrics Section on Hematology/Oncology and Committee on Genetics suggests an alternative for children who are allergic to penicillin: "Erythromycin prophylaxis may be used as an alternative for children with suspected or proven penicillin allergy" (American Academy of Pediatrics 2002 and 2016). Therefore, we have included a broader definition of antibiotoic prophylaxis to avoid an underestimation of the proportion of children with sickle cell anemia who are protected against pneumococcal infection.
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	N/A

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

References

American Academy of Pediatrics Section on Hematology/Oncology and Committee on Genetics. Health supervision for children with sickle cell disease. *Pediatrics* 2002; 109(3):526-535.

American Academy of Pediatrics. AAP publications reaffirmed or retired. Pediatrics 2016; 137(5):e20160592.

Bjornson AB, Falletta JM, Verter JI, Buchanan GR, Miller ST, Pegelow CH, et al. Serotype-specific immunoglobulin G antibody responses to pneumococcal vaccine in children with sickle cell anemia: effects of continued penicillin prophylaxis. *J Pediatr*1996;129(6):828-835.

Falletta JM, Woods GM, Verter JI, Buchanan GR, Pegelow CH, Iyer RV, et al. Discontinuing penicillin prophylaxis in children with sickle cell anemia. Prophylactic Penicillin Study II. *J Pediatr* 1995; 127(5):685-690.

Gaston MH, Verter JI, Woods G, Pegelow C, Kelleher J, Presbury G, et al. Prophylaxis with oral penicillin in children with sickle cell anemia. A randomized trial. *N Engl J Med* 1986; 314(25):1593-1599.

Gaston MH, Verter J. Sickle cell anaemia trial. *Stat Med* 1990; **9**(1-2):45-51.

John AB, Ramlal A, Jackson H, Maude GH, Sharma AW, Serjeant GR. Prevention of pneumococcal infection in children with homozygous sickle cell disease. *Br Med J (Clin Res Ed)* 1984; 288(6430):1567-1570.

Nkouwap I, Diara JP, Noyon I, Etienne-Julan M, Merault L. Is there any alternative to oral penicillin in antibioprophylaxis for children with sickle cell disease? [French] Y a-t-il une alternative a la penicilline orale dans l'antibioprophylaxie chez les enfants drepanocytaires? *Med Mal Infect*. 1999; 29(2):111-116.

Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273(5):408–412.

Woods GM, Jorgensen JH, Waclawiw MA, Reid C, Wang W, Pegelow CH, et al. Influence of penicillin prophylaxis on antimicrobial resistance in nasopharyngeal S. pneumoniae among children with sickle cell anemia. The Ancillary Nasopharyngeal Culture Study of Prophylactic Penicillin Study II. *J Pediatr Hematol Oncol* 1997; **19**(4):327-333.

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

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

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

QMETRIC_SCA_AntibioticProphylaxis_NQF_EvidenceAttachment-636166164555645567.docx

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

Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

1b. Performance Gap

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

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

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*
IF a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)
IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.
Children with SCA are at an increased risk of infection compared with children without the disorder. Daily receipt of antibiotic prophylaxis substantially reduces the risk of infection among these high-risk children. Prior studies indicate that appropriate antibiotic prophylaxis rates are low; however, these reports are limited in their generalizability, as they are usually focused on a single healthcare provider or registry. This measure establishes a claims-based method for identifying appropriate antibiotic prophylaxis among larger and broader populations of children with SCA. The performance scores calculated through this measure will identify areas in need of improvement of antibiotic prophylaxis among children with SCA.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*
Rates of appropriate antibiotic prophylaxis dispensed for at least 300 days within the measurement year for children with sickle cell anemia in six states (2005-2010).

Florida
(Year: Numerator/Denominator = Rate)
2005: 27/176 = 15.34%
2006: 21/169 = 12.43%
2007: 23/125 = 18.40%
2008: 18/155 = 11.61%
2009: 30/238 = 12.61%
2010: 44/282 = 15.60%

Illinois
(Year: Numerator / Denominator = Rate)
2005: 19/55 = 34.55%
2006: 24/77 = 31.17%
2007: 26/72 = 36.11%

2008: 21/72 = 29.17%
2009: 22/96 = 22.92%
2010: 13/75 = 17.33%

Louisiana
(Year: Numerator / Denominator = Rate)
2005: 16/104 = 15.38%
2006: 12/89 = 13.48%
2007: 36/126 = 28.57%
2008: 32/128 = 25.00%
2009: 27/120 = 22.50%
2010: 19/120 = 15.83%

Michigan
(Year: Numerator / Denominator = Rate)
2005: 9/42 = 21.43%
2006: 8/56 = 14.29%
2007: 10/70 = 14.29%
2008: 13/73 = 17.81%
2009: 7/96 = 7.29%
2010: 20/95 = 21.05%

South Carolina
(Year: Numerator / Denominator = Rate)
2005: 12/51 = 23.53%
2006: 9/52 = 17.31%
2007: 12/60 = 20.00%
2008: 2/34 = 5.88%
2009: 2/35 = 5.71%
2010: 11/41 = 26.83%

Texas
(Year: Numerator / Denominator = Rate)
2005: 28/84 = 33.33%
2006: 20/77 = 25.97%
2007: 30/100 = 30.00%
2008: 26/122 = 21.31%
2009: 31/123 = 25.20%
2010: 38/136 = 27.94%

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

1b.4. **Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.** *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*
The dataset used for performance scores is Medicaid; therefore, there are no disparities identified by insurance or socioeconomic status. However, the majority of children with sickle cell disease (approximately 80%) have been enrolled in Medicaid at some point in time. In addition, the majority (over 80%) of children with sickle cell disease enrolled in Medicaid are enrolled for at least one continuous year (Reeves SL, Fullerton HJ, Cohn LM, et al. Missed opportunities for transcranial Doppler screening among children with sickle cell disease. Clin Pediatr 2016; 55(12): 1093-1099). Further, these statistics are among a broader population of

children with sickle cell disease; we would expect the number of children with SCA who are continuously enrolled in Medicaid to be even higher considering that SCA is the most clinically significant variant of sickle cell disease.

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

Several studies have pointed to disparities in prophylactic medication use among patients with public versus private insurance. In a study of children with sickle cell disease on Medicaid in Washington state and Tennessee, 10.3% of patients with public insurance received no penicillin or macrolide antibiotic or trimethoprim-sulfamethoxazole during a 365-day period while only 21.5% received more than 270 days of medication in a year. Median duration of prescriptions was 10 days (Sox et al., 2003). In a 10-year retrospective cohort study of 407 infants enrolled in the Tennessee Medicaid program, 60% of infants with sickle cell disease did not have recommended prophylactic antibiotic prescriptions filled within the recommended period (i.e., the first 12 weeks of life) (Warren et al., 2010). A study assessing compliance with penicillin prophylaxis for sickle cell disease found that adherence was significantly greater in patients with private versus public insurance (17/28 [61%] vs. 33/90 [37%], respectively) (Teach et al., 1998.) Clearly, however, significant room for improvement remains, despite type of insurance.

References

Sox CM, Cooper WO, Koepsell TD, DiGiuseppe DL, Christakis DA. Provision of pneumococcal prophylaxis for publicly insured children with sickle cell disease. JAMA 2003; 290(8):1057-1061.

Teach SJ, Lillis KA, Grossi M. Compliance with penicillin prophylaxis in patients with sickle cell disease. Arch Pediatr Adolesc Med 1998; 152(3):274-278.

Warren MD, Arbogast PG, Dudley JA, et al. Adherence to prophylactic antibiotic guidelines among Medicaid infants with sickle cell disease. Arch Pediatr Adolesc Med 2010; 164(3):298-299.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

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

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
http://chear.org/sites/default/files/stories/pdfs/qmetric_sca_antibioticprophylaxsis_nqf_specification.pdf

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

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

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

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

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator is the number of children ages 3 months to 5 years old with SCA (Hb SS) who were dispensed appropriate antibiotic prophylaxis for at least 300 days within the measurement year.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Target population (children with SCA): Children with SCA (Hb SS) are identified through the presence of at least three separate healthcare encounters related to Hb SS within the measurement year. These encounters are identified through either ICD-9 or ICD-10 codes. Children ages 3 months to 5 years are included within the target population (i.e., must not have a 6th birthday within the measurement year). Children must be continuously enrolled within the health plan in which claims are available and must have no other form of health insurance for the entire measurement year.

Cases from target population with target process (appropriate antibiotic prophylaxis dispensed for at least 300 days within the calendar year): Antibiotic prophylaxis is defined as at least 300 days covered within the measurement year, which is the summed total of the number of days' supply of antibiotics dispensed within the measurement year (see National Drug Codes (NDC) table attached in S.2b.).

NOTE: Although NHLBI guidelines specifically recommend penicillin for antibiotic prophylaxis, some children may have or be suspected to have penicillin sensitivity. The American Academy of Pediatrics Section on Hematology/Oncology and Committee on Genetics suggests an alternative for children who are allergic to penicillin: "Erythromycin prophylaxis may be used as an alternative for children with suspected or proven penicillin allergy" (Citation: American Academy of Pediatrics Section on Hematology/Oncology and Committee on Genetics (Pediatrics 2002; 109(3):526-535; Reaffirmed in 2016). Therefore, we have included a broader definition of antibiotic prophylaxis than penicillin in this measure. This is intended to avoid underestimation of the proportion of children with SCA who are protected against pneumococcal infection.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator is the number of children ages 3 months to 5 years with SCA (Hb SS) within the measurement year.

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Children with SCA (Hb SS) are identified through the presence of at least three separate healthcare encounters related to Hb SS within the measurement year. Hb SS-related healthcare encounters are identified through either ICD-9 or ICD-10 codes (See specification in S.1). Children ages 3 months to 5 years are included within the target population (i.e., must not have a 6th

birthday within the measurement year). Children must be continuously enrolled within the health plan in which claims are available and must have no other form of health insurance for the entire measurement year.

Note: Children with SCA are included starting at 3 months of age to account for any lag in identification and confirmation of the sickle cell disease status of the child.

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
There are no denominator exclusions.

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

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

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
No risk adjustment or risk stratification
If other:

S.12. **Type of score:**
Rate/proportion
If other:

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

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

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
This measure does not involve sampling; all SCA cases meeting the inclusion criteria are used in the calculation.

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
NA

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Claims (Only)

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
NA

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

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

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Other
If other: Any setting represented with prescription medication claims data

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

2. Validity – See attached Measure Testing Submission Form
QMETRIC_SCA_AntibioticProphylaxis_NQF_TestingAttachment.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.3 For maintenance of endorsement
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

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

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

What were the statistical results of the analyses used to select risk factors?

Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)
No - This measure is not risk-adjusted

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: **Antibiotic Prophylaxis Among Children with Sickle Cell Anemia**

Date of Submission: 12/6/2016

Type of Measure:

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

Instructions

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

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

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

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

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]

AND

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

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

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

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

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

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

Notes

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

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

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

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

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

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

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

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

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

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☒ abstracted from paper record

☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☒ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☒ other: Michigan Newborn Screening

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

- Michigan Medicaid administrative claims data provided by the Michigan Department of Health and Human Services (MDHHS)
- Medicaid Analytic eXtract (MAX) administrative claims data for six state Medicaid programs provided by the Centers for Medicare & Medicaid Services (CMS)

Other data used for testing (not existing datasets):

- 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
- Michigan Newborn Screening (NBS) Results

1.3. What are the dates of the data used in testing? Michigan Medicaid data 2007-2011; MAX data: 2005-2010; CHM, HMC, and UMHS medical record data: 2012; Michigan NBS: 1987-2010

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

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

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

- The Michigan Medicaid data consisted of all Medicaid claims for Medicaid enrollees within the state of Michigan (2007-2011).
- The MAX data consisted of all Medicaid claims reported to CMS for Medicaid enrollees within six state Medicaid programs with moderate to high prevalence of sickle cell anemia: Florida, Illinois, Louisiana, Michigan, South Carolina, and Texas (2005-2010).
- The medical record data were obtained from three hospitals: CHM, HMC, and UMHS (2012). These three large medical centers are located in urban areas in Michigan reflective of the residence of the vast majority 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.

- The Michigan NBS data consisted of all births within the state of Michigan (1987-2010).

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

- The Michigan Medicaid data from 2007 to 2010 (used for performance score validation) was a complete census of all children ages 3 months to 5 years with sickle cell anemia who met eligibility criteria within each year (Table 1). The population was equally divided between sexes; approximately 98% were black.

Table 1: Number of children ages 3 months to 5 years with sickle cell anemia enrolled in Michigan Medicaid, 2007-2010

2007	2008	2009	2010
114	118	149	141

- The Michigan Medicaid data from 2010 and 2011 provided a complete census of all children ages 1-18 years (used to validate the denominator element) with at least one sickle cell disease (SCD)-related administrative claim, continuously enrolled annually within Michigan Medicaid in 2010 and/or 2011, with a newborn screening result available. This included 938 children in 2010 and 924 children in 2011. The population was equally divided between sexes; approximately 75% were black and the average age was approximately 10 years.

- The MAX data included all children ages 3 months to 5 years with sickle cell anemia who met eligibility criteria within each year for Medicaid claims reported by selected states (Table 2). The population was equally divided between sexes; approximately 98% were black.

Table 2: Number of children enrolled in Medicaid, ages 3 months to 5 years, with sickle cell anemia, MAX data by state, 2005-2010

State	2005	2006	2007	2008	2009	2010
Florida	176	169	125	155	238	282
Illinois	55	77	72	72	96	75
Louisiana	104	89	126	128	120	120
Michigan	42	56	70	73	96	95
South Carolina	51	52	60	34	35	41
Texas	84	77	100	122	123	136

- A sample of abstracted medical records from 34 children with sickle cell anemia ages 3 months to 5 years who were enrolled in Michigan Medicaid was drawn at three sickle cell centers in Michigan (CHM, HMC, UMHS) during 2012.
- The Michigan NBS data included all children born in the state of Michigan from 1987-2010 with a positive and confirmed screening result that had at least one sickle cell disease-related claim and who were continuously enrolled for 12 months in Michigan Medicaid in either the year 2010 or 2011.

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

- Reliability testing data: MAX
- Validity testing data: Michigan Medicaid, MAX, Michigan NBS, and medical records

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

The data do not include patient-level sociodemographic (SDS) variables; however, all children included in the data were enrolled in Medicaid.

2a2. RELIABILITY TESTING

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

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

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

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

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

The reliability of MAX data to evaluate appropriate antibiotic prophylaxis 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. This 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).

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

State-specific reliability results for appropriate antibiotic prophylaxis among children with sickle cell anemia are detailed in Table 3. These results show that the reliability based on signal-to-noise analysis ranged from 0.83 to 0.96, with a median of 0.89.

Table 3. State-specific reliability for measure

State	Numerator	Denominator	Reliability
Florida	163	1145	0.9592
Illinois	125	447	0.8475
Louisiana	142	687	0.9130
Michigan	67	432	0.8919
South Carolina	48	273	0.8251
Texas	173	642	0.8909
Median (range)			89.14% (82.51-95.92)

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

State-specific reliability is very good; observed reliability was consistently greater than 0.80. 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 (RAND Corporation, TR-653-NCQA, 2009). The median reliability observed across states was 0.89 (range: 0.83-0.96), which is consistent with a high degree of reliability.

2b2. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

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

Critical Data Elements

Numerator: The accuracy of administrative claims in identifying antibiotic prophylaxis was assessed through comparison with 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 measure specification criteria. Each medical record was assessed for the presence of a prescription for antibiotics and compared with administrative claims to determine if the prescription was filled and the difference between the date of prescription and the prescription fill. 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.

Denominator: 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.

Conversion of ICD-9 to ICD-10

The goal of ICD-9 to ICD-10 conversion was to convert this measure to a new code set, fully consistent with the intent of the original measure. All ICD-9 diagnosis codes were converted to the corresponding ICD-10 codes using the CMS 2015 diagnosis code General Equivalence Mappings (GEMs) and diagnosis code description files (accessed on August 26, 2015); these mapping files were created by CMS. The target ICD-9 codes were converted to ICD-10 using the GEM file and manually reviewed for consistency using the diagnosis code descriptions for the source ICD-9 and converted ICD-10 codes. In addition, the resultant ICD-10 codes were back-translated to ICD-9 to verify the accuracy of the coding. Source files from CMS were acquired from these files:

1. ICD 9 to 10 diagnosis GEM -*2015_I9gem.txt*
   https://www.cms.gov/Medicare/Coding/ICD10/2015-ICD-10-CM-and-GEMs.html

2. ICD 10 to 9 diagnosis GEM - 2015_10gem.txt
   https://www.cms.gov/Medicare/Coding/ICD10/2015-ICD-10-CM-and-GEMs.html
3. ICD 9 description file *CMS32_DESC_SHORT_DX.txt*
   https://www.cms.gov/Medicare/Coding/ICD9ProviderDiagnosticCodes/codes.html
4. ICD 10 description file - *icd10cm_order_2015.txt*
   https://www.cms.gov/Medicare/Coding/ICD10/2015-ICD-10-CM-and-GEMs.html

The ICD-9 code 282.61 (Hb SS disease without crisis) mapped to the ICD-10 code of D57.1 (sickle-cell disease without crisis). This ICD-10 code was not included in the measure specification, as it is not specific to sickle cell anemia (Hb SS). The ICD-9 code 282.62 (Hb SS disease with crisis) mapped to ICD-10 D57.00 (Hb SS disease with crisis, unspecified) and was included in the specification. Subsequent verification using the GEMs indicated that ICD-10 codes D57.01 (Hb SS disease with acute chest syndrome) and D57.02 (Hb SS disease with splenic sequestration) were also appropriate to include in the measure specification to identify the study population (denominator).

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) with the gold standard of Michigan Medicaid data (obtained directly from Michigan's claims data warehouse) for the same time period (2007-2010). 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 antibiotic prophylaxis 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.

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 (QMETRIC). The QMETRIC 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 QMETRIC 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 antibiotic prophylaxis, the panel weighed evidence to determine if the performance 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.

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

Critical Data Elements

Numerator: For this comparison, 34 children with sickle cell anemia who were enrolled within Michigan Medicaid were successfully matched with their Michigan Medicaid administrative claims data. Among these children, 25 cases (76%) had a Medicaid administrative claim for an antibiotic prescription filled within 29 days of the prescription date in the medical record. An additional seven cases (21%) had an antibiotic prescription filled between 30 but less than 90 days following the prescription date. One case (3%) had an antibiotic prescription claim filled 90 or more days after the prescription date in the medical record. Eighteen charts were also chosen for calculation of inter-rater reliability; the two trained abstractors had 100% agreement with each other for abstracting antibiotic prescriptions from the medical records, resulting in a kappa of 1.00.

Denominator: For this comparison, 865 children met eligibility criteria in 2010 (at least one SCD-related claim ages 1-18, continuous enrollment in Michigan Medicaid in 2010, a NBS result available); 836 children met eligibility criteria in 2011. In 2010, a case definition of three Hb SS 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 (Table 4). 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.

Table 4. Accuracy of case definition of at least 1, 2 and 3 Hb SS claims within a year to identify children with sickle cell anemia compared with the gold standard of newborn screening

Algorithm	Area under the ROC Curve	# True Positives	# False Positives	# True Negatives	# False Negatives	Sensitivity	Specificity	PPV	NPV
Results - 2010									
≥1 Hb SS Claim	0.50	409	456	0	0	100.0%	0.0%	47.3%	NA
≥2 Hb SS Claims	0.82	391	144	312	18	95.6%	68.4%	73.1%	94.5%
≥3 Hb SS Claims	0.86	374	91	365	35	91.4%	80.0%	80.4%	91.3%
Results - 2011									
≥1 Hb SS Claim	0.50	397	439	0	0	100.0%	0.0%	47.5%	NA
≥2 Hb SS Claims	0.79	377	163	276	20	95.0%	62.9%	69.8%	93.2%
≥3 Hb SS Claims	0.87	363	97	342	34	91.4%	77.9%	78.9%	91.0%

Empirical Validity Testing of Performance Measure

The comparison of rates of appropriate antibiotic prophylaxis from the gold standard of Michigan Medicaid data compared with MAX data can be seen in Table 5. This illustrates that the number of children who were dispensed at least 300 days of antibiotics among children with sickle cell anemia ranged from 14 to 23 in the claims acquired directly from the Medicaid data warehouse, versus a range of 7 to 20 from MAX data for the same time period.

Table 5: Comparison of appropriate antibiotic prophylaxis by source of Medicaid claims data for the state of Michigan, 2007-2010

Source	Rate Components	2007	2008	2009	2010
Michigan Medicaid data	Numerator	14	23	22	22
	Denominator	114	118	149	141
	Percentage	12.28%	19.49%	14.77%	15.60%

MAX data	Numerator	10	13	7	20
	Denominator	70	73	96	95
	Percentage	14.29%	17.81%	7.29%	21.05%

Figure 1 illustrates the performance scores observed between the Michigan Medicaid data from the state warehouse and MAX data from CMS for each overlapping year noted, respectively: 12.3% versus 14.3% (2007); 19.5% versus 17.8% (2008); 14.8% versus 7.3% (2009), and 15.6% versus 21.1% (2010).

Figure 1: Comparison of appropriate antibiotic prophylaxis by source of Medicaid claims data for the state of Michigan, 2007-2010

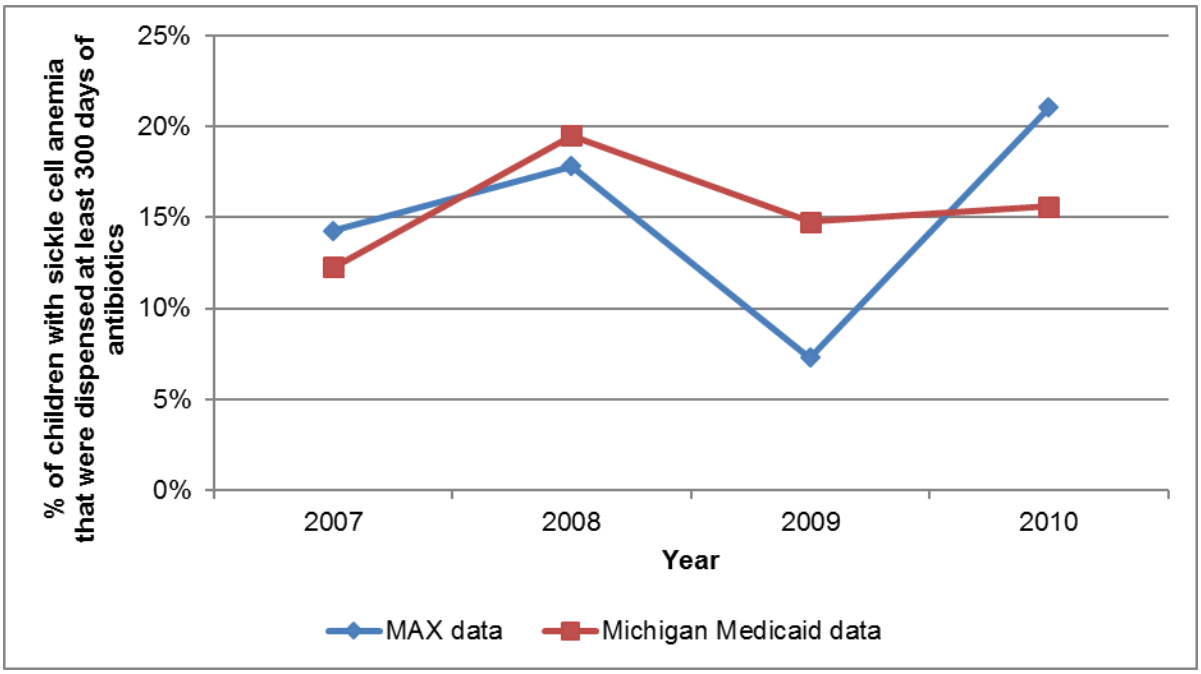


Table 6 reports the z-scores and p-values from the two-sample z-tests comparing the proportion of children who were dispensed at least 300 days of antibiotic prophylaxis each year between Michigan Medicaid and MAX data.

Table 6: Comparison of performance score by source of Medicaid claims data, Michigan

	2007	2008	2009	2010
z-score	0.3921	-0.2890	-1.7678	1.0735
p-value	0.6965	0.7718	0.0767	0.2846

Face Validity of Performance Measure Score

The QMETRIC 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 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.

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

Critical Data Elements

Numerator: The majority (33/34, 97%) of prescribed antibiotics in the medical record were reflected in administrative claims. Further, 30/34 (88%) of those who were prescribed an antibiotic in the medical record were dispensed their antibiotic within 30 days of the prescription. Given this evidence, we believe the validity of administrative claims in assessing antibiotic prescriptions is very high.

Denominator: A sensitivity of over 90% and a specificity of approximately 80%, as well as the reliability across years, allow us to conclude that the denominator is valid for accurately identifying children with sickle cell anemia within administrative claims. These results indicate that the case definition used has a very high ability to correctly identify true cases and a somewhat lower ability to distinguish false positives. However, other less stringent case definitions resulted in substantially more misclassification than the chosen definition of at least three Hb SS claims within the measurement year.

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 antibiotic prophylaxis was assessed for the same state (Michigan) from these two data sources for the same time period (2007-2010), no differences in rates were observed (all p-values >0.05). Therefore, our results suggest that compared with Michigan Medicaid data, MAX data is highly valid.

Face Validity of Performance Measure Score

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

2b3. EXCLUSIONS ANALYSIS

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

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

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

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

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

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

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

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

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

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

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

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

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

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

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

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

2b4.9. Results of Risk Stratification Analysis:

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

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

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

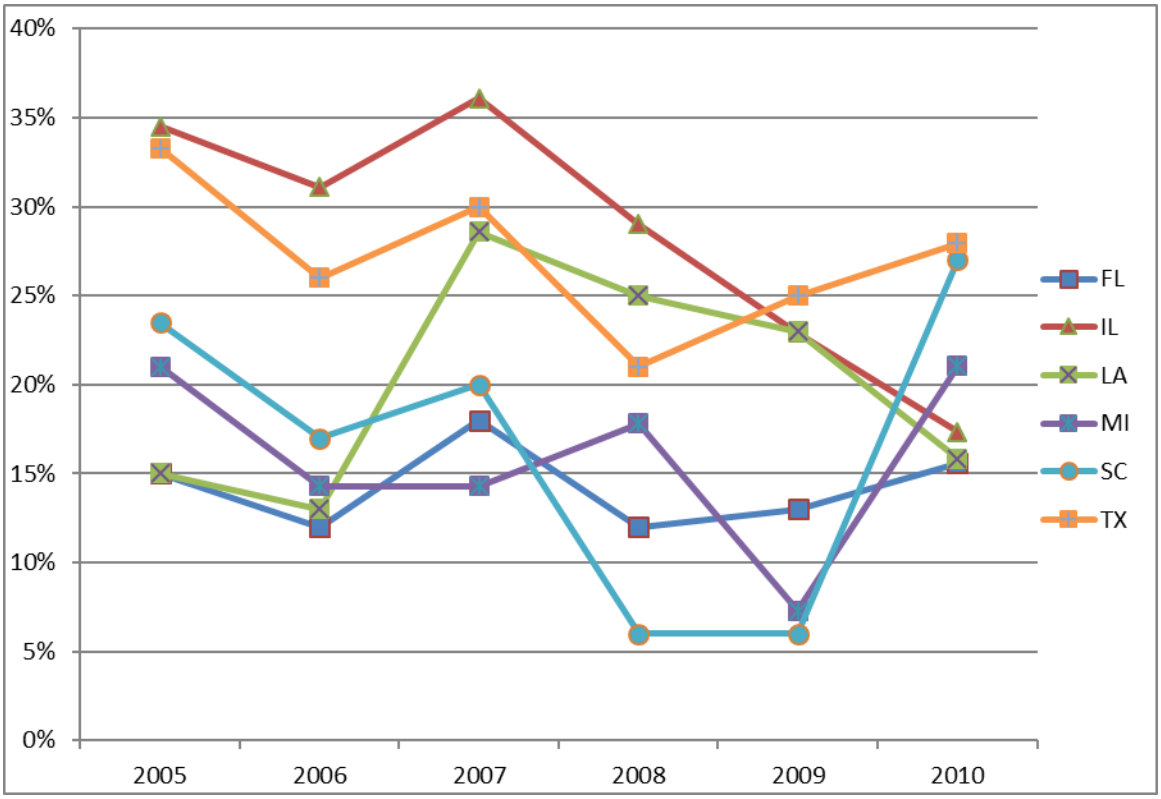
1b)

Using the MAX data, the proportion of children dispensed at least 300 days of antibiotic prophylaxis 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 at least 300 days of antibiotic prophylaxis, 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. For all models, regression diagnostics were performed to assess normality of error variances; healthcare utilization and demographics were adjusted for within the model. In addition, this model was assessed to determine if differences at the state health plan level could be assessed using this measure.

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

The proportion of children who were dispensed at least 300 days of antibiotics ranged from 3.23% to 36.11% (Figure 2).

Figure 2. Provision of antibiotic prescriptions, 300-day rate, for children with sickle cell anemia using MAX data, 2005-2010



Compared with 2005, children did not have higher odds of being dispensed at least 300 days of antibiotic prophylaxis (Table 7). However, rates did vary by state (Table 8).

Table 7. Odds of being dispensed at least 300 days of antibiotic prophylaxis among children with sickle cell anemia enrolled in six state Medicaid programs by year using MAX data, 2005-2010

Year	Odds Ratio	95% Confidence Interval
2005	Reference	Reference
2006	0.74	0.48, 1.13
2007	0.94	0.62, 1.41
2008	0.72	0.47, 1.09

2009	0.42	0.26, 0.69
2010	0.83	0.55, 1.28

Table 8. Odds of being dispensed at least 300 days of antibiotic prophylaxis among children with sickle cell anemia enrolled in Medicaid programs by state using MAX data, 2005-2010

State		
FL	Reference	Reference
IL	3.12	1.95, 4.97
LA	1.73	1.01, 2.77
MI	1.60	0.94, 2.70
SC	1.33	0.71, 2.50
TX	3.37	2.19, 5.19

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

Although this measure is successfully able to distinguish differences in performance across states and years, the rates of appropriate antibiotic prophylaxis are not increasing over time.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

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

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

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

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

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

We tested the completeness and accuracy of Medicaid claims for sickle cell patients seen at three major medical centers in Michigan. For this comparison, children with sickle cell anemia who were enrolled within Michigan Medicaid were matched with their Michigan Medicaid administrative claims data (n=34).

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

Of these children, the vast majority of prescribed antibiotics in the medical record were reflected in Medicaid claims (33/34, 97%). We found that the preponderance of those who were prescribed an antibiotic in the medical record had a corresponding Medicaid claims for a dispensed antibiotic within 30 days of the prescription (76%). An additional seven cases (21%) had an antibiotic prescription filled between 30 but less than 90 days following the prescription date. We noted only one case (3%) that did not have an antibiotic prescription claim filled within 90 days of a prescription date.

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

Given this evidence, we believe that missing data is unlikely to bias our performance results, particularly as any missing data would be expected to be non-differential across entities.

3. Feasibility

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

3a. Byproduct of Care Processes

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

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims)
If other:

3b. Electronic Sources

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

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

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

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

3c. Data Collection Strategy

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

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.
This measure was tested using Medicaid administrative claims data. The primary information needed for this measure includes basic demographics, diagnosis codes, and procedure codes and dates. These data are widely available, although obtaining them often requires a restricted-use data agreement. For multiple-state comparisons, MAX data are available from CMS. When the measure is used at the single-state level, state health departments can use their own Medicaid data.

QMETRIC testing determined that this measure, which is intended to be used with administrative claims data systems, works well in that environment. The measure was also tested for reliability and validity using medical chart data, from both paper and electronic health record (EHR) sources. Continuing advances in the development and implementation of EHRs may establish the feasibility of regularly implementing this measure with data supplied by EHRs.

While QMETRIC testing efforts support the feasibility of implementing this measure, the testing process demonstrated the technical challenges involved in identifying appropriate SCD cases from very large administrative claims files such as MAX data.

This measure was also tested using administrative claims data acquired directly from the state of Michigan. Acquisition of data directly from states requires the cooperation of those jurisdictions, as well as modification of the statistical programming code developed for MAX files to correctly function using the unique structure of the data files obtained from each state.

This measure was also validated using information acquired from medical records from Children's Hospital of Michigan (where each patient had both a paper record and an EHR), and from Hurley Medical Center and the University of Michigan Health System (both institutions had only EHR information). However, EHR systems did not support structured queries of the relevant data; chart abstraction was also necessary. The abstraction process demonstrated that, while the required data for this measure can be obtained from both types of systems, EHRs could provide greater efficiency if structured queries were possible.

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

4. Usability and Use

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

4a. Accountability and Transparency
4a. Accountability and Transparency

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

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

Specific Plan for Use	Current Use (for current use provide URL)
Public Health/Disease Surveillance Quality Improvement (external benchmarking to organizations) Quality Improvement (Internal to the specific organization)	

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

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

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
Under our recently awarded PQMP grant, the New York Medicaid program has partnered with QMETRIC to formally begin field testing and implementation of this measure in 2017 at the state, plan, and health system levels.

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

implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.)
NA

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

4b. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
NA

4c. Unintended Consequences
The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**
No unintended negative consequences to individuals or populations were identified during testing.

4c.2. **Please explain any unexpected benefits from implementation of this measure.**
This measure was considered for inclusion in the CMS Core Set. It will be reconsidered if NQF endorsement is conferred.

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
NA

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
NA

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.
NA

4d2.2. **Summarize the feedback obtained from those being measured.**
NA

4d2.3. **Summarize the feedback obtained from other users**
NA

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
NA

5. Comparison to Related or Competing Measures

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

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

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

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

5a. Harmonization of Related Measures
The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

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

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

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

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

Appendix

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

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** QMETRIC - University of Michigan
Co.2 **Point of Contact:** Sarah, Reeves, sleasure@med.umich.edu, 734-615-9755-
Co.3 **Measure Developer if different from Measure Steward:** QMETRIC – University of Michigan
Co.4 **Point of Contact:** Gary, Freed, gfreed@med.umich.edu, 734-615-3139-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**

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

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

The QMETRIC 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

QMETRIC 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)
Sarah Reeves, PhD, MPH, Research Investigator of Pediatrics, School of Medicine, University of Michigan, Ann Arbor, MI

Measure Developer/Steward Updates and Ongoing Maintenance

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

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

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

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

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:** This work was 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 and contract number HHSP233201600166A. 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. The Measures can be reproduced and distributed, without modification, for noncommercial purposes, e.g., use by health care providers in connection with their practices. Commercial use is defined as the sale, license, or distribution of

the Measures for commercial gain, or incorporation of the Measures into a product or service that is sold, licensed or distributed for commercial gain. Commercial uses of the measures require a license agreement between the user and the Quality Measurement, Evaluation, Testing, Review and Implementation Consortium (QMETRIC) at the University of Michigan (U-M). Neither QMETRIC/U-M nor their members shall be responsible for any use of the Measures. QMETRIC/U-M makes no representations, warranties or endorsement about the quality of any organization or physician that uses or reports performance measures, and QMETRIC/U-M has no liability to anyone who relies on such measures. The QMETRIC performance measures and specifications are not clinical guidelines and do not establish a standard of medical care.

This statement is signed by Gary L. Freed, MD, MPH, who, as the principal investigator of QMETRIC, is authorized to act for any holder of copyright on the submitted measure.

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, QMETRIC
Child Health and Evaluation Research (CHEAR) Unit
Division of General Pediatrics
University of Michigan Hospital and Health Systems
Ann Arbor, MI 48109-5456

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

NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Brief Measure Information

NQF #: 3189
Measure Title: Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: Visits per 100 Child-years
Measure Steward: University Hospitals Cleveland Medical Center
Brief Description of Measure: This measure estimates the rate of emergency department visits for children ages 2 – 21 who are being managed for identifiable asthma, using specified definitions. The measure is reported in visits per 100 child-years.
Developer Rationale: Asthma is a critical problem with racial and ethnic disparities and varies by urbanicity. Adherence to the National Asthma Education and Prevention Programs (NAEPP) Guidelines improves outcomes. [1-32]. We have elsewhere provided other articles, studies, and summaries of evidence to document that ED visits and hospitalizations are typically outcome measures of choice when assessing asthma control.

ED visits for asthma in children are common and expensive. They may result from poor quality of care delivered (failure to adhere to guidelines) as well as from insufficient access to primary care. Asthma is the leading diagnosis leading to urgent care/emergent care provided in emergency departments for children. It is among the most common chronic diseases in children and expenses for asthma care are in the billions of dollars annually. Further, CMS and AHRQ assigned us this measure. In addition to data and citations provided, the team has analyzed 2007 and 2011 waves of the National Survey of Children's Health and confirmed that this parent reported measure both identified a high prevalence of asthma nationwide and significant consequences in terms of parent reported child health for children who have asthma.

Our analysis of National Survey of Children's Health [33] data (NSCH, 2011/12), estimates that 10.3 million children in the U.S. have been told that they have asthma. Of these children 7.6 million live in more urban areas that are characterized as metropolitan statistical areas (MSAs), an asthma prevalence rate of 15.4%. Table 1 shows that asthma is very consequential for health.

Table 1. Impacts of Asthma for Children Age 2-17, NSCH 2011/12
Parent/caregiver reports child's health status is excellent or very good

	2 - 5 years	6 – 11 years	12 – 17 years	Total
All Children living in Metropolitan Statistical Areas				
Asthma	59.8 %	69.6 %	74.3 %	70.1 %
No asthma	87.8 %	85.3 %	85.1 %	85.9 %
Overall	84.9 %	82.8 %	83.1 %	83.4 %
Difference	-28.0 %	-15.7 %	-10.8 %	-15.8 %
Children living in MSAs with Asthma				
All Children	59.8 %	69.6 %	74.3 %	70.1 %
Black or Latino	52.1 %	64.1 %	66.4 %	62.9 %
Not Black/Latino	66.5 %	74.6 %	80.4 %	76.1 %
Difference	-14.4 %	-10.5 %	-14.0 %	-13.2 %

We find overall a 15.8% drop in the proportion of parents who report their child's health as very good or excellent among those who have asthma, and almost twice that in younger children. Because 2 of our networks are in the greater NYC area, these data highlight children who live in more urban areas. Outside of urban areas both prevalence and gap between those with and without asthma are slightly higher (each ~17%). Effective delivery of guideline-based care can reduce the gap and decrease consequences of uncontrolled asthma, such as emergency room use and hospitalizations; better asthma care is beneficial and

needed across the spectrum of children and primary care settings.[34-40] We find compelling evidence that the failure to effectively deliver guideline-based care contributes significantly to the lower health ratings for children with asthma, including for the 3.4 million urban Black and Hispanic children (age 2-17 years) with asthma. About 60% of these children are low income and have public insurance. We further are persuaded by evidence that quality of life and the quality of asthma management are associated specifically with such factors as family satisfaction with the nature of shared decision making.[41]

Citations for data demonstrating high priority

1. PCORI. PCORI Funding Annoouncement: Treatment Options for Afircan Americans and Hispanics/Latinos with Uncontrolled Asthma. 2013 [cited 2013 September 18]; Available from: http://pcori.org/assets/2013/06/PCORI-Asthma-PFA-061813.pdf.
2. Marcano-Belisario, J., Greenfield G, Huckvale K, Gunn LH, Car J, Apps for asthma self-management: a systematic assessment of content and tools. Cochrane Database Syst. Rev., 2012(8).
3. Health, O.o.M. Asthma and African Americans. [Fact Sheet]. 2012 [cited 2013 August 28]; Available from: http://minorityhealth.hhs.gov/templates/content.aspx?ID=6170.
4. Health, O.o.M. Asthma and Hispanic Americans. [Fact Sheet]. 2012 [cited 2013 August 28]; Available from: http://minorityhealth.hhs.gov/templates/content.aspx?ID=6173.
5. Wennergren, G., Strannegard I, Asthma hospitalizations continue to decrease in schoolchildren but hospitalization rates for wheezing illnesses remain high in young children. Acta Paediatr, 2002. 91(11): p. 1239-1245.
6. Wisnivesky, J., Lorenzo J, Lyn-Cook R, et al., Barriers to adherence to asthma management guidelines among inner-city primary care providers. Annals of Allergy, Asthma & Immunology, 2008. 101(3): p. 264-270.
7. DiSantostefano, R., Davis K, Yancey S, Crim C, Ecologic analysis of asthma-related events and dispensing of inhaled corticosteroid- and salmeterol-containing products. Ann Allergy Asthma Immunol, 2008. 100(6): p. 558-565.
8. Crocker, D., Kinyota S, et al., Effectiveness of home-based, multi-trigger, multicomponent interventions with an environmental focus for reducing asthma morbidity: a community guide systematic review. Am J Prev Med, 2011. 41(2): p. S5-S32.
9. Gustafson, D., Wise M, et al., The effects of combining Web-based eHealth with telephone nurse case management for pediatric asthma control: a randomized controlled trial. J Med Internet Res., 2012. 14(4): p. e101.
10. Program, T.N.A.E.a.P., Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. J Allergy Clin Immunol, 2007. 120(5 Suppl): p. S94-138.
11. Celano, M., Holsey CN, Kobrynski LJ, Home-based family intervention for low-income children with asthma: a randomized controlled pilot study. J Fam Psychol., 2012. 26(2): p. 171-178.
12. Zuniga, G., Kirk S, et al., The impact of asthma health education for parents of children attending head start centers. J Community Health, 2012. 37(6): p. 1296-1300.
13. Blaakman, S., Tremblay PJ, Halterman JS, Fagnano M, Borrelli B, Implementation of a community-based secondhand smoke reduction intervention for caregivers of urban children with asthma: process evaluation, successes and challenges. Health Educ Res, 2013. 28(1): p. 141-152.
14. Martin, M., Catrambone CD, et al., Improving asthma self-efficacy: developing and testing a pilot community-based asthma intervention for African American adults. J Allergy Clin Immunol, 2009. 123(1): p. 153-159.
15. Seid, M., D´Amico E, Varni JW, et al., The In Vivo Adherence Intervention For at Risk Adolescents With Asthma: Report of a Randomized Pilot Trial. Journal of Pediatric Psychology, 2012. 37(4): p. 390-403.
16. Edwards, J., INSPIRE curriculum delivered in a faith-based setting. Fam Community Health., 2010. 33(2): p. 117-122.PRINCIPAL INVESTIGATOR
17. Press, V., Pappalardo AA, et al., Interventions to improve outcomes for minority adults with asthma: a systematic review. J Gen Intern Med, 2012. 27(8): p. 1001-1015.
18. DeJongh, T., Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R, Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst. Rev., 2012.
19. James, T., Fine M, Monitoring Asthma Control Using Claims Data And Patient-Reported Outcomes Measures. P.T., 2008. 33(8): p. 454-466.
20. Bender, B., Overcoming barriers to nonadherence in asthma treatment. J Allergy Clin Immunol, 2002. 109(6): p. 554-559.
21. Okelo, S., Eakin M, et al., The Pediatric Asthma Control and Communication Instrument asthma questionnaire: for use in diverse children of all ages. J Allergy Clin Immunol, 2013. 132(1): p. 55-62.
22. Benavides, S., Rodriquez JC, Maniscalco-Feichtl M, Pharmacist involvement in improving asthma outcomes in various healthcare settings: 1997 to present. Ann Pharmacother. , 2009. 43(1): p. 85-97.
23. Drotar, D., Physician behavior in the care of pediatric chronic illness: association with health outcomes and treatment adherence. J Dev Behav Pediatr., 2009. 30(33): p. 246-254.
24. Weinstein, A., The potential of asthma adherence management to enhance asthma guidelines. Ann Allergy Asthma Immunol, 2011. 106(4): p. 283-291.

25. Agency, E.P. President's Task Force on Environmental Health Risks and Safety Risks to Children:Coordinated Federal Action Plan to Reduce Racial and Ethnic Asthma Disparities. 2012 [cited 2013 August 28]; Available from: http://www.epa.gov/

Numerator Statement: The numerator estimates the number of emergency department (ED) visits for asthma among children being managed for asthma. Since most hospitalizations for asthma are from the ED and many ED visits that result in hospitalization are not captured in encounter data, a numerator event may be identified either as an ED visit or as a hospitalization.

Denominator Statement: The denominator represents the person time experience among eligible children with identifiable asthma. Assessment of eligibility is determined for each child monthly. The total number of child months experienced is summed and divided by 1200 to achieve the units of 100 child years for the denominator.

Assessing eligibility for the denominator requires 2 years of data, the reporting year and the 12 month period before the reporting year. (See Appendix 1, Figure 1)

Denominator Exclusions: Children with specified concurrent or pre-existing diagnosis and children who have not been consecutively enrolled in the reporting plan for at least three months, as specified in the details section.

Measure Type: Outcome

Data Source: Claims (Only), Claims (Other)

Level of Analysis: Health Plan, Population : Community, County or City, Population : Regional and State

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

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary

This measure was previously reviewed by the Pulmonary Standing Committee (March 2016) as NQF 2794 and has been revised and resubmitted. In the measure's evidence and testing forms, content submitted previously is in black; new information is in blue in the measure testing and evidence forms.

The developer provides the following rationale for this outcome (plan or population) measure:

- Accessible, high-quality primary care reduces the need for emergency department (ED) visits by decreasing the number of children who have acute breakthrough episodes requiring the ED.
- Accessible, high-quality primary care reduces the need for ED visits by decreasing the number of children who come to the ED for asthma care better performed in the office setting.
- ED visits/hospitalizations are undesirable outcomes that can be reduced by better primary care management.
- A systematic review of the body of evidence is not required for outcome measures.
- The evidence for this measure is based on clinical practice guidelines for asthma control from the National Heart and Lung and Blood Institutes (NHLBI) (2007): "As a general rule, patients with well-controlled asthma should have: ... no emergency department visits; no hospital stays ...". Grade C = Nonrandomized trials and observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.
- Although not required per NQF guidance for outcome measures, the developer also provided three systematic reviews: Interventions to Modify Health Care Provider Adherence to Asthma Guidelines; Cochran Database of Systematic Reviews: Intermittent versus daily inhaled corticosteroids for persistent asthma in children and adults (Review); Quality of Care for Childhood Asthma: Estimating Impact and Implications.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Guidance from the Evidence Algorithm

Assesses health outcome (Box 1) → Relationship between outcome and healthcare action (Box 2) → Pass

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

(Previous review by Pulmonary Committee: Pass - Evidence: Y-21; N-0)

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

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

The developer reports:

- ED visits for asthma are common, may be reduced through improved primary care or community-based interventions, and demonstrate disparities.
- NHLBI NAELPP guideline provides a description of clinical evidence of gaps.
- The developer reports 10.3 million children have asthma and that asthma has a significant impact on health, with an overall a 15.8% drop in the proportion of parents who report their child's health as very good or excellent among those who have asthma, and almost twice that in younger children.
- The developer reports overall rate of ED visits for asthma in NY State Medicaid Managed Care in 2012 is 20.65 per 100 child-years.
  - By age stratum the rates are 47.4 visits per 100 child-years for children 2 to 4 years, 26.0 for children 5 to 11 years, 22.7 for adolescents 12 to 18 years, and 34.1 for adolescents 19 to 21 years.
  - There are differences in performance by race, urbancity, and quartile of poverty.
  - The developer provides additional data demonstrating expected seasonal variations in performance rates.

Disparities

The developer reports:

- Asthma is a critical problem with racial and ethnic disparities and varies by urbanicity. The developer's analysis of National Survey of Children's Health data (NSCH, 2011/12), estimates that 10.3 million children in the United States have been told that they have asthma. Of these children, 7.6 million live in more urban areas that are characterized as metropolitan statistical areas (MSAs), with an asthma prevalence rate of 15.4%.
- The developer reports that, on a yearly and a monthly basis, differences exist in performance by age, urbanicity, race/ethnicity, and level of poverty. Additionally, it identifies disparities in cross tabulations—e.g., the performance rate for children 2 to 4 years in large metropolitan areas is 52.6 visits per 100 child-years compared to those in small metropolitan areas with 26.2 visits per child year, in micropolitan areas with 18.3 visits/100 child-years, and in rural areas with 12.3 visits per 100 child-years.
- The developer reports racial and ethnic differences were notable:
  - For children ages 2 to 4 years, the rate in non-Hispanic Whites was 18.4 visits per 100 child-years, in Asians 19.3 visits per 100 child-years, in Hispanics 53.9 child-years, and in non-Hispanic Blacks 74 visits per 100 child-years.
  - The disparities regardless of age were Black, 41.99 visits per 100 child-years; White, 14.79 visits per 100 child-years; Hispanic, 31.91 visit per 100 child-years.

Question for the Committee:

Does the Committee believe there is a gap in care that warrants a national performance measure?

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

Previous review by Pulmonary Committee: Pass - Performance Gap: H-6; M-14; L-1; I-0

Committee pre-evaluation comments

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

- Asthma ED visits and hospitalizations are widely viewed as health outcomes for children. The developers provide sufficient links data to conclude that a healthcare process (controller regimens) can affect the outcome.
- I have concerns about the lower age limit of 2 years old. Not seeing substantive guidelines to substantiate inclusion of the 2 year old age group. Where in NHLBI guidelines does it support this for inclusion.
- This is an outcome measures and the developers cite evidence that ED visits are a signal of quality and that interventions by providers can reduce rates of ED visits in this patient population
- The developers provide ample evidence of high rates of ED use and hospitalization, a significant fraction of which is preventable. The marked differences by race/ethnicity, urban residence, and SES are further strong evidence of performance gaps. I would rate this as "high".
- It demonstrates disparities, however, it is not clear how complete New York Medicaid data are with regards to race. What percent of Medicaid enrollee race is incomplete?
- "Distribution of ED visit rates by measured entity (health plan and county level) are shown in item 2b5.2 rather than here. The results shown do demonstrate variation in quality across measured entities. However they do show substantial racial and ethnic disparities in outcome rates (no in performance at the plan or provider level). However this demonstrates that the measure could be used to illuminate and potentially reduce disparities at the population level.
- The developers appear to provide differences in the performance on the measures across plans in the measure reliability section. Although they do not show the distribution of scores they show that performance is different."

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability specifications

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

Data source(s):

- Administrative claims

Specifications:

- The level of analysis is plan or population
- The developer defines the numerator as: *The numerator estimates the number of emergency department (ED) visits for asthma among children being managed for asthma. To enhance validity, a numerator event may be identified either as an ED visit or as a hospitalization.*
- The denominator for this measure is: *The person time experience among eligible children with identifiable asthma. Assessment of eligibility is determined for each child monthly. The total number of child months experienced is summed and divided by 1200 to achieve the units of 100 child years.*
- The exclusions for the measure are: *Children with specified concurrent or pre-existing diagnosis and children who have not been consecutively enrolled in the reporting plan for at least three months, as specified in the details section.*
- The developer states "If pharmacy data are not available, the measure should be reported with notation that pharmacy data were not used for the assessment of eligibility. This avoids eliminating from the measure those facilities with no link to pharmacies. Our testing reveals that only a small proportion of patients are excluded by not including pharmacy data to establish eligibility." However, specifics on the amount of missing data or the proportion of patients with missing data were not submitted.
- The numerator and denominator details include the CPT and ICD-9 codes; ICD-10 codes are included in an attachment.
- This outcome measure is not risk adjusted.
- The calculation algorithm is stated in S.14.
- There is no sampling for this measure.

Questions for the Committee:

- *Are the definitions and codes for "managed for identifiable asthma" and "asthma related medication" appropriate? Are they specific enough so they can be reliably collected by different parties?*
- *The developer notes the goal of the measure is to assess how many children are visiting the ED for asthma treatment. Is the numerator specifications 'either an ED visit **or hospitalization**' appropriate?*
- *Are all the data elements clearly defined? Are all appropriate codes included?*
- *Is it likely this measure can be consistently implemented?*
- *Is the potential variability in access to/inclusion of pharmacy data a concern?*

2a2. Reliability Testing, Testing attachment

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

SUMMARY OF TESTING

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

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

Method(s) of reliability testing

- Per NQF guidance separate reliability testing is not required if data element-level validity testing is performed, however, the developer also stated it performed zero inflated poisson (ZIP) approach. This approach is generally not considered a demonstration of reliability, but rather a demonstration of differences in performance.

Results of reliability testing

- The current information provided is not sufficient to demonstrate reliability. NQF staff have requested more information from the developer, but it is not yet available.

Questions for the Committee:

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

Guidance from the Reliability Algorithm: Precise specifications (Box 1)→ Empirical reliability testing (Box 2) → Empirical validity testing of patient level data (Box 3) → Insufficient

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

RATIONALE: The current information is not sufficient to demonstrate reliability.

Previous review by Pulmonary Committee: Pass - Reliability: H-2; M-17; L-2; I-0

2b. Validity

2b1. Validity: Specifications

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

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

- The goal of the measure is to assess how many children are visiting the ED for asthma treatment. According to the developer, ED visits for asthma are a function of a sick child who needs to be seen; poor access to high-quality primary care; or poor quality management of a chronic condition. The rate should be low, but not zero. The numerator of children with undesirable visits and a denominator of children with identified asthma are consistent with this evidence.

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

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

SUMMARY OF TESTING

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

Method of validity testing of the measure score:

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

Validity testing method:

- The developer states: "Assessment of the capacity to identify the eligible population and qualifying events was performed in NY State Medicaid data in both 2011 and 2012 reporting years. For MCO analysis we analyzed both with the 18 plans that had 900 or more children contributing to the denominator and with the 20 plans that contributed at least 1000 months of person time to the denominator."
- For both the numerator and denominator, the developer relies on literature to support its conclusion of the validity of administrative data elements to identify children who are being managed with identifiable asthma. Per NQF policy:
  - Prior evidence of validity of data elements can be used, including published data, provided it includes the same data elements; uses the same data type; and is conducted on an appropriate sample (i.e., representative, adequate numbers, etc.)
  - The developer attests that the data elements match those assessed in the literature.
- However, the developer did not cite the full range of literature (as cited in the other submission, #2816) to support this measure.
- The developer used NY State Medicaid Managed Care claims data for its analyses.
- The developer states it used face validity, but its face validity assessments involve construct of the measure. NQF specifically requires that face validity be at the measure score level and that the assessment be that the measure ***score*** can distinguish good from poor quality.
- The developer also cites score-level validity testing of two previously-endorsed asthma measures as evidence of data-element level validity. However, this does not meet NQF's requirements for demonstration of data element validity.
  - The developer states there is "nearly complete overlap of the denominator codes for this measure and there is overlap of the denominator elements." The developer states that where codes differ, they were specific decisions by its expert panel.
  - The information provided is insufficient for denominator validity as the specific differences and how they do or do not affect the validity are not described.

Validity testing results:

The developer reports:

- "The literature also supports the use of claims data to identify the presence of asthma. We use administrative data to identify the age of the child, various stratification variables and the presence of asthma, as well as the presence of an asthma ED visit or hospitalization. These are routinely used to support billing by CMS, Medicaid, and private insurers and are routinely used in quality measurement. There is moderate agreement (kappa = 0.45 – 0.50) when comparing administrative data regarding the presence of constructs such as recent asthma attacks, use of asthma medications, attack or medication, attack and medication, using 1 year of administrative claims data. The agreement improves from 0.55 to 0.60 when using two years of data as this measure does."

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

The developer provides the following information:

- Denominator exclusions include: Children with concurrent or pre-existing: Chronic Obstructive Pulmonary Disease (COPD) diagnosis, Cystic Fibrosis diagnosis, or Emphysema diagnosis. Children who were not consecutively enrolled in the reporting plan for three consecutive months ending in the reporting month are excluded.
- There are no numerator exclusions.
- <=2.5% potentially eligible children were excluded by clinical diagnoses.
- Exclusions are clinical and represent construct validity rather than statistical considerations. Longer continuous enrollment requirements would harm the validity of the measure since more children with multiple diagnosis would have been excluded. The 3-month continuous enrollment requirement is provided so that the child is under the management of the health plan, which is the accountable unit.

Question for the Committee:

- *Are any patients or patient groups inappropriately excluded from the measure?*

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

Conceptual rationale for SDS factors included? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

The developer provides the following information:

- "Specifications for this measure require stratification and reporting by age group only and also within age group by race/ethnicity." The submission indicates that the stratification is informational, not to control for patient characteristics.
- The developer states biological data and national guidelines do not support risk adjustment to control for patient characteristics.

Conceptual analysis of the need for SDS adjustment:

- The developer notes additional stratification variables are optional (e.g., rurality/urbanicity and county level of poverty), but may be required by the accountability entity or reported by the reporting entity. According to the developer, "risk adjustment is not critical for interpreting the results or for validity, but ... stratification is informative to help to promote like to like comparisons and allow for plans to demonstrate how they do on specified subgroups. Such voluntary stratification specified in the measure helps to mitigate against the potential for misinterpretation and unintended consequences."
- The developer acknowledges the association of the risk factors with performance on the measure, but states risk adjustment is not justified by such differences as "either acceptable or unmodifiable by health care," and posits that evidence exists that primary care, adherence to guidelines, and other interventions can reduce or eliminate the impact of the risk factors.

Empirical analysis of the SDS factors:

- The developer also found that urban counties perform differently than rural counties and the size of the county impacts performance.

- The developer also found differences in performance by race: ED utilization of Blacks is significantly different from Whites ($p<0.01$); ED utilization of Hispanics is significantly different from Whites ($p<0.01$); ED utilization of Blacks and Hispanics are significantly different from one another ($p<0.01$).

Question for the Committee:

- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*

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

- The developer used NY State Medicaid Managed Care claims data, comparing 18 plans that included at least 900 children in the denominator, which "yielded statistically significant differences among plans and among counties, whether or not we controlled for age group and/or race and ethnicity and/or urbanicity. This is true also when we analyzed stratified by age group."
  - The developer notes that partially managed plans had a statistically higher rate than fully managed HMO plans. Among all 18, the mean rate is 15.7, with a standard deviation of 6.0.
- The developer states that "Poisson Regression analyses indicate significant differences by health plans and by counties, whether or not controlling for age group and race/ethnicity."
  - The ZIP models showed that even after controlling for age, urban and rural counties performed differently, as did suburban and rural counties vs small urban counties. There were also statistically significant differences in ED usage between whites and blacks, blacks and Hispanics, and whites and Hispanics.
- The developer analyzed meaningful differences related to the stratification subpopulations (e.g., within a state), reporting it performed Chi-square analysis and t-statistics. The developer states "differences between major groups were statistically significant", $p<0.05$.
- The developer states the measure is sensitive enough to detect meaningful differences as observed within a population across counties and between counties and NYC.
- The developer also notes that "Comparing to a randomly selected index plan, 14 of 17 plans had statistically significant differences in performance with the median and modal p-value being <0.001. Non-significant plans' p-values=0.08, 0.16 and 0.88."

Questions for the Committee:

- *Does this measure identify meaningful differences about quality at the health plan level?*
- *Does this measure identify meaningful differences about quality at the population level?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

The developer notes the following:

- Since administrative claims are used, the extent of missing data is expected to be low. There were a total of three children in our analysis of children with identifiable asthma who dropped out of the analyses because of any missing data element.
- The developer states, "Our analyses found that the absence of pharmacy data would reduce only slightly (as we recall, less than 1%) the number of children identified as having identifiable asthma; no specific data are provided in the submission. This finding became apparent during alpha testing of our specifications and was incorporated into our specifications as a permissive allowance when pharmacy data were not available. We have not located the original analysis and hope for the NY State team to replicate the analysis by the time of the Committee meeting."

- The developer also states that “Systems unable to integrate pharmacy data into the eligibility analysis would have a minimally higher risk population than those with pharmacy claims. The specifics of the definitions and the limited impact of pharmacy claims on eligibility combine to make the expected impact of this on the rate of ED visits to almost zero. They are included in the identification of denominator because our expert panel directed us to do so…. for this measure, pharmacy data is used only to complement other utilization data when determining eligibility.”

Guidance from the Validity Algorithm
Measure specifications consistent with evidence (Box 1)→ All potential threats to validity are empirically assessed (Box 2)→ Insufficient

The highest possible rating is INSUFFICIENT.

Preliminary rating for validity: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**
RATIONALE: The current submission is insufficient, until further information is submitted regarding risk stratification.
Previous review by Pulmonary Committee: Consensus Not Reached - Validity: H-0; M-10; L-11; I-0

Committee pre-evaluation comments

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

- The general constructs are reasonable. Specific considerations include:
  - The combination of ED visits and hospital admissions has reasonable face validity, and the data presented support their combination.
  - The codes used in the numerator appear reasonable.
  - The eligibility screen for the denominator is complex. The inclusion of events in the prior year PLUS events in the months of the measurement year prior to the measurement month is unusual. The developers do not really explain why a child in January of the measurement year would have the prior 12 months to achieve the criteria for ""identifiable asthma"" while a child in November of the measurement year would have 22 mos (12 in the prior year, and the 10 months before November in the measurement year). (If I have misunderstood the approach, my apologies.) I am not sure that it introduces bias that would affect the comparison of health care entities on this measure.
  - I am puzzled by the inclusion of children >5 on the basis of 1 ambulatory visit plus an ED visit in the measurement month- only for children in this age group.
  - The inclusion of visits with ""bronchitis"" is unusual as a criterion for identifiable asthma. I do not see data on how including or excluding these diagnoses would change the denominator in a given health plan or population.
  - The expert panel process is described well. No rationale is given for why short acting beat-agonists are excluded in the denominator of ""identifiable asthma"". "
- Would seek additional clarification from the developers as to whether all inclusion criteria are assess within the past month. For example, a hospitalizations with a principal of secondary discharge diagnosis of asthma counts only if it happened in the last month? If so this measure cohort is clearly focused on a population of patients with active asthma symptoms or recent exacerbations. This seems logical as these are the patients at greatest risk of using the ED and for whom providers should be actively engaged in prevention or early management to avoid ED visits.
- "The reliability testing appears reasonable. The developers tested this measure in a large Medicaid claims environment. The use of zero inflated poisson models makes intuitive sense as described, but this does not seem to be acceptable by the NQF. This should be reviewed by a statistician. Evidence on the reliability of the data elements is also specified in terms of how data in medical charts are used to reliably code the diagnoses. There is not a robust discussion of the impact of including secondary vs. primary diagnosis codes in the numerator definition. I would like to understand better why the ZIP method is not seen as contributing to reliability leading to the preliminary assessment of the NQF staff as ""Inadequate"".
- Some reliability testing is provided under the validity items. However, they do not present a true test re-test validity test of the measure score at the plan or county level. I would encourage them to provide this to the committee
- "Validity testing was done only in one state's Medicaid data. While it would have been better if this had been across several states, or included commercial plans as well, this is a fairly large and diverse population in which to test this measure.

- The validity of the data elements is well-supported by the literature, though the developers do not really provide evidence for how changes in specifications (not using short-acting beta agonists) might affect validity. I agree that the developers do not explicitly make a claim of face validity of the score as a measure of quality, but it is implied and I think a reasonable conclusion. The findings by race/ethnicity and urban residence are all very consistent with studies in the literature, further supporting the validity."
- My primary concern is over the short enrollment period requirement of 3 months for inclusion in the denominator. According to the authors, this cohort accounts for 20% of the denominator. 3 months of enrollment (when you factor in that it takes around 2 full weeks for insurance card to arrive to home of new enrollee; and another 30 days or so for PCP selection; it does not seem valid to consider 3 months of enrollment as sufficient time for at least one physician visit to occur (which would either be a sick or well-child visit), to include asthma diagnosis, and expect asthma to be managed well-enough to prevent ED visit. seems like an unreasonable expectation. Have the authors compared their outcome data with and without the 20% who were enrolled for only 3 consecutive months?
- Measure score validity testing is not provided. Developers should be encouraged to include a systematic rating of the validity of the measure score from their TEP if available.
- The developers describe the impact of lack of pharmacy data for determining eligibility for a measurement month. The developers provide an argument for why this will not affect the measure very much.
- The authors did not discuss if they allowed for a three month run-out of the administrative data to ensure data completeness for claims. How much time elapsed after the end of the reporting period before the data were collected from the state or health plans? Need to hear more about data completeness and what percent of claims were complete (paid).
- "Exclusions appropriate with minor impact on cohort. Risk adjustment -- I agree with the developer's approach to stratify the measure by age group and to provide stratified results to illuminate disparities. The exclusion of children with chronic lung disease is likely sufficient and additional risk-adjustment not needed. However, I would have preferred to see some analysis to ensure that there is no biased distribution of chronically ill asthma patients across health plans. Meaningful differences: adequately demonstrated by developers. Comparability and missingness do not seem to be relevant for this measure.

Criterion 3. Feasibility

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

- All data elements are in defined fields in electronic claims.
- The developer reports there are no fees.

Question for the Committee:

o *Are the required data elements routinely generated and used during care delivery?*

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

Previous review by Pulmonary Committee: Pass - Feasibility: H-15; M-6; L-0; I-0

Committee pre-evaluation comments

Criteria 3: Feasibility

- The feasibility of the basic measure is adequate, since it relies on data that is available in claims data. As the developers note, some of the stratification variables (e.g. race/ethnicity) may not be available and may vary in how reliable they in different systems. The developers argue that the measure will push health systems to improve data collection and reporting. This is possible, but not assured.
- Highly feasible as dependent on information available in hospital billing/claims. However, stratification would depend on accuracy of demographic information at the plan or state level which may be poor and highly variable. I would have like to have see some more detailed description of quality and quality assessment of these data.

Criterion 4: Usability and Use

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

Current uses of the measure
Publicly reported? ☐ **Yes** ☒ **No**

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

Accountability program details

- This is a new measure so current use is not required.
- The developer is working on specific plans for dissemination and use.
- The developer is discussing application and use of this measure with New York State Medicaid.
- The developer plans for the measure to be used for an accountability application within three years of NQF endorsement and public reporting within six years of initial endorsement.

Improvement results

- As a new measure, the developer does not present progress on improvement.
- The developer states a variety of stakeholders would benefit from this measure, e.g., clinicians, health systems, state and healthcare agencies, researchers, etc.

Unexpected findings (positive or negative) during implementation
N/A

Potential harms

- The developer reports no unintended negative consequences to individual or populations during testing.
- The developer reports possible unintended/negative consequences and recommends against the following:
  - Comparing individual health care professionals.
  - A single hospital comparison because this measure is intended to measure system performance not the hospital performance.
  - Measuring anything other than large practices or integrated delivery systems that own their own risk and manage inpatient and outpatient care or that have access to all payer data sources.

Vetting of the measure

- None provided

Feedback:

- No feedback provided on QPS. MAP has not reviewed this measure for inclusion in any federal program.

***Question for the Committee**:*

- *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
Previous review by Pulmonary Committee: Pass - Usability and Use: H-4; M-11; L-5; I-1
Not recommended on Overall Vote for Endorsement: Y-3; N-15

Committee pre-evaluation comments

Criteria 4: Usability and Use

- The measure is not currently in use. The basic measure should be useable, but use of some of the stratifications may have additional barriers at least in the short to medium term.

Criterion 5: Related and Competing Measures

Related or competing measures

The developer did not include information on any of the related or competing measures. However, NQF staff identified the following measures that may be related and/or competing.

- 0047: Asthma: Pharmacologic Therapy for Persistent Asthma
- 0728: Asthma Admission Rate (PDI 14)
- 1800: Asthma Medication Ratio
- 2414: Pediatric Lower Respiratory Infection Readmission Measure
- 2816: Appropriateness of Emergency Department Visits for Children and Adolescents with Identifiable Asthma (*submitted by the same developer for review in this project)*

Harmonization

No information provided

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

RATIONALE IF NOT ELIGIBLE: This measure is not eligible for Endorsement+ because it has face validity testing only and has not been vetted by those being measured or other users.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: Visits per 100 Child-years

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

Date of Submission: 12/14/2016

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*

- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

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

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

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

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

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

Outcome

☒ Health outcome: ED asthma visits for children with identifiable asthma

☐Patient-reported outcome (PRO): Click here to name the PRO

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

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

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Emergency department (ED) visits are often linked to the management of a child's asthma. Emergency Department Asthma was the topic assigned to CAPQuaM for measurement. ED visits for children with asthma is an outcome measure of intrinsic value. It represents utilization of an expensive service and constitutes a burden on children and their families. Two literature reviews as well as focused reviews that we have done to supplement the extensive

review of the literature confirms the importance of an integrated approach to managing the health care of children with asthma. There is abundant evidence that ED visits are common, may be reduced through improved primary care or community-based interventions, and demonstrate disparities (1-11, 12-19). Asthma is generally recognized to be an ambulatory care sensitive condition. Nonetheless, we perceive and our panel articulated that the rate for ED visits ought not to be 0. So while in general a lower rate represents preferable care, too low a rate could indicate insufficient access to emergency room services. Our overarching conceptual framework that extends beyond this measure is shown in the evidence form.

~~Our measure benefits from a formal development process, CAPQuaM's 360 degree method, which is described in more detail in the measure testing form. The measure and its specifications result from a formal development process for this measure incorporated stakeholder input including a parent focus group, meeting with The Mount Sinai Pediatrics Department's Parent Advisory Council, interviews with primary care clinicians and ED physicians, the CAPQuaM's multidisciplinary scientific team, which includes investigators, a steering committee and a senior advisory board of nationally prominent figures. The measure also benefits from a national multidisciplinary Expert Panel which utilized a RAND type modified Delphi method to guide our specifications.~~

When epidemiologists describe how frequently something occurs the preferred measure is typically an incidence density, or rate. In contrast to a risk or proportion, the incidence density has as its denominator a measure of the extent of potential exposure in the population, expressed in people-years. This measure represents an advance in the measurement of healthcare performance for children: it incorporates this formulation both to enhance its interpretation (because it has a specific epidemiological meaning) and to limit distortion if sick children move in or out of eligibility for the measure. (20)

Further clinical evidence of gaps are demonstrated in the description by NHLBI's NAELPP guideline, cited in the evidence form, Schatz and colleagues study describing the relationship between asthma control and asthma exacerbations in managed care (21), and Fuhlbrigge et al's confirmation that medications can work to reduce ED visits for asthma but are used sub optimally (22). When children with asthma experience adequate management of chronic conditions and have access to coordinated care, a reduction in hospital rates is likely to occur. (23) Children who are linked to continuous care utilize less overall care, including ED care. (23)

The following diagram presents an overview of how CAPQuaM conceptualizes asthma ED visits for children with asthma.

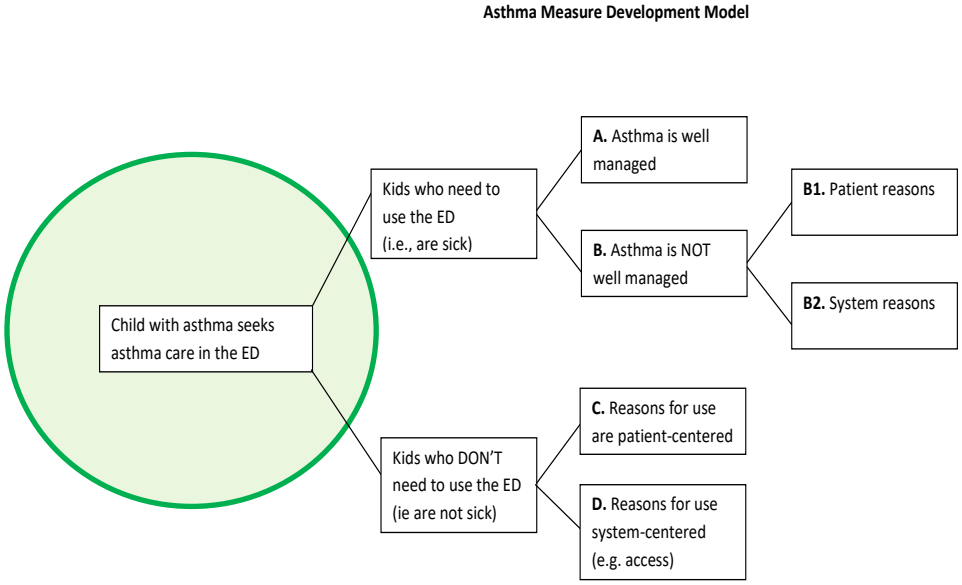


Figure Notes: The green circle highlights that this measure identifies which children who present to the emergency room should be considered to represent an ED visit for a child who is being managed for identifiable asthma.

References

1. Adams, R.J., B.J. Smith, and R.E. Ruffin, Factors associated with hospital admissions and repeat emergency department visits for adults with asthma. Thorax, 2000. 55(7): p. 566-73.
2. Bahadori, K., et al., Economic burden of asthma: a systematic review. BMC Pulm Med, 2009. 9: p. 24.
3. Weiss, K.B., P.J. Gergen, and T.A. Hodgson, An economic evaluation of asthma in the United States. N Engl J Med, 1992. 326(13): p. 862-6.
4. Coventry, J.A., M.S. Weston, and P.M. Collins, Emergency room encounters of pediatric patients with asthma: cost comparisons with other treatment settings. JAmbul Care Manage, 1996. 19(2): p. 9-21.
5. Association, A.L. American Lung Association Fact Sheet on Asthma and Children. October 2012 [cited 2013 May 20]; Available from: http://www.lung.org/lung-disease/asthma/resources/facts-and-figures/asthmachildren-fact-sheet.html#5.
6. Fuhrman, C., et al., Hospitalizations for asthma in children are linked to undertreatment and insufficient asthma education. J Asthma, 2011. 48(6): p. 565-71.
7. Sawicki, G.S., et al., Uncontrolled asthma in a commercially insured population from 2002 to 2007: trends, predictors, and costs. J Asthma, 2010. 47(5): p. 574-80.
8. Manice, M., Exploring the relationship between parental shared decision-making practices and acute asthma exacerbations among children age 0-17. 2013, Icahn School of Medicine at Mount Sinai: New York, NY.
9. Cerdan, N.S., et al., Asthma severity in children and the quality of life of their parents. Appl Nurs Res, 2012. 25(3): p. 131-7.
10. Fiese, B., et al., Family climate of routine asthma care: associating perceived burden and mother-child interaction patterns to child well-being. Fam Process, 2008. 47(1): p. 63-79.
11. Okelo, S.O., et al., Emotional quality-of-life and outcomes in adolescents with asthma. J Pediatr, 2004. 145(4): p. 523-9.
12. Talreja, N., et al., Modifiable factors associated with severe asthma exacerbations in urban patients. Ann Allergy Asthma Immunol, 2012. 109(2): p.128-32.
13. Auger, K.A., et al., Medical home quality and readmission risk for children hospitalized with asthma exacerbations. Pediatrics, 2013. 131(1): p. 64-70.
14. Ducharme, F.M., et al., Written action plan in pediatric emergency room improves asthma prescribing, adherence, and control. Am J Respir Crit Care Med, 2011.183(2): p. 195-203.
15. Farber, H.J., Optimizing maintenance therapy in pediatric asthma. Curr Opin Pulm Med, 2010. 16(1): p. 25-30.
16. Smith, S.R., D.B. Wakefield, and M.M. Cloutier, Relationship between pediatric primary provider visits and acute asthma ED visits. Pediatr Pulmonol, 2007.42(11): p. 1041-7.
17. Self, T.H., et al., Reducing emergency department visits and hospitalizations in African American and Hispanic patients with asthma: a 15-year review. J Asthma, 2005. 42(10): p. 807-12.
18. Lara, M., et al., Reducing quality-of-care disparities in childhood asthma: La Red de Asma Infantil intervention in San Juan, Puerto Rico. Pediatrics, 2013. 131 Suppl 1: p. S26-37.
19. Oraka, E., et al., Racial and ethnic disparities in current asthma and emergency department visits: findings from the national health interview survey, 2001-2010. J Asthma, 2013. 50(5): p. 488-96.
20. Rothman, K.J., Modern Epidemiology. 3rd edition ed. 2008, Philadelphia, PA: Lippincott, Williams and Wilkins.
21. Schatz M, Zeiger RS, Yang ST, et al. “Relationship of asthma control to asthma exacerbations using surrogate markers within a managed care database.” Am J Managed Care. 16(5):327-333, 2010.
22. Fuhlbrigge A, Carey VJ, Adams RJ. “Evaluation of asthma prescription measures and health system performance based on emergency department utilization.” Med Care 42(5):1-7, 2004.
23. Cooley W, McAllister J, Sherrieb K, Kuhlthau K. *Improved outcomes associated with medical home implementation in pediatric primary care.* Pediatrics, 2009. **124**(1): p. 358-364.

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

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

An abundant literature supports both that emergency department visits and hospitalizations are considered undesirable outcomes for asthma and that at a population level these undesirable outcomes can be reduced by better clinical management, including medication management, the use of asthma action plans, and effective and continuous primary care, among other things. Asthma is considered to be an ambulatory care sensitive condition further reinforcing the consensus in the field that utilization of ED visits and/or hospitalizations are generally (at the population level) preventable when managed in an ambulatory setting within our current knowledge.

1. Accessible high quality primary care reduces the need for ED visits by decreasing the number of children who have acute breakthrough episodes requiring the ED or inpatient setting.
2. Accessible high quality primary care reduces the need for ED visits by decreasing the number of children who come to the ED for asthma care better performed in the office setting.

As ED visits and/or hospitalizations can represent significant cost for families and for the system, asthma is the single most prevalent diagnosis leading to ED visits for children in the USA, urgent asthma visits to the ED can be disruptive for families, and both ED visits and hospitalizations are not free of iatrogenic and nosocomial risk, these outcomes have intrinsic importance.

To support this rationale there are four specific systematic reviews of the evidence cited. Detailed information on each can be found in the Appendix Table 6-9. Highlights are below:

(1) National Heart, Lung, and Blook Institute, National Institutes of Health (NHLBI/NIH) Asthma Guideline 2007

Quick Reference Guide: Asthma control focuses on two domains: 1) reducing impairment --- the frequency and intensity of symptoms… and 2) reducing risk – the likelihood of future asthma attacks… [later described as "prevent exacerbations]

At the population level ED visits and hospitalizations represent failures of asthma control.

Asthma Guidelines:

- **Following science-based guidelines *works***
- Not only do they have the potential to improve a patient's *quality* of life; they can potentially *save a life*.

National asthma guidelines have been updated: In 2007, the National Asthma Education and Prevention Program (NAEPP), coordinated by the National Heart, Lung, and Blood Institute (NHLBI), released its third set of clinical practice guidelines for asthma. The Expert Panel Report 3—Guidelines for the Diagnosis and Management of Asthma (EPR-3) reflects the latest scientific advances in asthma drawn from a systematic review of the published medical literature by an NAEPP-convened expert panel. It describes a range of generally accepted best-practice approaches for making clinical decisions about asthma care.

The EPR-3 emphasizes the importance of asthma control and focuses on two domains—current impairment and future risk—by which to assess asthma severity (for initiating therapy) and asthma control (for ongoing monitoring). EPR-3 also includes an expanded section on childhood asthma (with an additional age group), new guidance on medications, new recommendations on patient education in settings beyond the physician's office, and new advice for controlling environmental exposures that can cause asthma symptoms.

Asthma can be controlled

Scientific evidence clearly shows that most people could control their asthma by following current asthma clinical practice guidelines. With proper care, people who have asthma can stay active, sleep through the night, and avoid having their lives disrupted by asthma attacks.

As a general rule, patients with well-controlled asthma should have:

- Few, if any, asthma symptoms.
- Few, if any, awakenings during the night caused by asthma symptoms.
- No need to take time off from school or work due to asthma.
- Few or no limits on full participation in physical activities.
- No emergency department visits.
- No hospital stays.
- Few or no side effects from asthma medicines.

KEYPOINTS: OVERVIEW OF MEASURES OF ASTHMA ASSESSMENT AND MONITORING (pg. 36)

- The functions of assessment and monitoring are closely linked to the concepts of severity, control, and responsiveness to treatment:
    - Severity: the intrinsic intensity of the disease process. Severity is measured most easily and directly in a patient not receiving long-term-control therapy.
    - Control: the degree to which the manifestations of asthma (symptoms, functional impairments, and risks of untoward events) are minimized and the goals of therapy are met.
    - Responsiveness: the ease with which asthma control is achieved by therapy.
- Both severity and control include the domains of current impairment and future risk:
    - Impairment: frequency and intensity of symptoms and functional limitations the patient is experiencing or has recently experienced
    - Risk: the likelihood of either asthma exacerbations, progressive decline in lung function (or, for children, reduced lung growth), or risk of adverse effects from medication

KEY DIFFERENCES FROM 1997 AND 2002 EXPERT PANEL (pg. 37) REPORTS

- The key elements of assessment and monitoring are refined to include the separate, but related, concepts of severity, control, and responsiveness to treatment. Classifying severity is emphasized for initiating therapy; assessing control is emphasized for monitoring and adjusting therapy. Asthma severity and control are defined in terms of two domains: impairment and risk.
- The distinction between the domains of impairment and risk for assessing asthma severity and control emphasizes the need to consider separately asthma's effects on quality of life and functional capacity on an ongoing basis (i.e., in the present) and the risks it presents for adverse events in the future, such as exacerbations and progressive loss of pulmonary function. These domains of asthma may respond differentially to treatment.

… p.38: An important point linking asthma severity, control, and responsiveness is that the goals are identical for all levels of baseline asthma severity. A patient who has severe persistent asthma compared to a patient who has mild persistent asthma, or a patient who is less responsive to therapy may require more intensive intervention to achieve well-controlled asthma; however, the goals are the same: in well-controlled asthma, the manifestations of asthma are minimized by therapeutic intervention.

… page 41 regarding identification asthma, one key factor is:

The Expert Panel recommends that the clinician trying to establish a diagnosis of asthma should determine that (EPR–2 1997):

- **Episodic symptoms of airflow obstruction are present.**

This is consistent with how we defined identifiable asthma…

Page 63: It is important to evaluate the frequency, rate of onset, severity, and causes of exacerbations…severe exacerbations leading to ED visits and hospitalizations (Adams et al. 2000; Eisner et al. 2001; Ford et al. 2001; Lieu et al. 1998).

(2) Interventions to Modify Health Care Provider Adherence to Asthma Guidelines: A Systematic Review

Demonstrates several tools are effective in enhancing the quality of care and reduce undesirable outcomes.

(3) Cochran Database of Systematic Reviews: Intermittent versus daily inhaled corticosteroids for persistent asthma in children and adults (Review)

Different approaches to treatment achieve different outcomes in children and adults (Daily achieves better asthma control than intermittent inhaled corticosteroids)

(4) Quality of Care for Childhood Asthma: Estimtating Impact and Implications

Identified multiple gaps in asthma care quality. Key outcomes identified include hospitalizations and emergency department visits. Identified large racial disparities in use of inhaled corticosteroids

In addition to the work cited above, we conducted a scoping review as follows:

We identified key constructs of asthma ED use measures for consideration. We created a table of these constructs in technical and lay language, and listed research questions for the review to answer. Our contractor (a national accrediting body experienced in measure development), prepared for us a literature review in 2 stages and we supplemented this with targeted reviews as needed to answer specific questions that arose during the measure development process.

The construct table (Appendix; pg 26) was used to guide the review and was the basis for the first round of review. Following the table, we include a list of questions for focused review (Appendix: pg 39) that guided round 2 of the review, which resulted in a detailed summary of 91 articles from the peer-reviewed literature. In addition to this review, the CAPQuaM scientific team conducted an ad hoc series of reviews to answer specific questions such as the reliability of administrative data to identify asthma, and the value of expert panels and the RAND/UCLA appropriateness method. The CAPQuaM degree 360 method starts with a topic area and the measures emerge during the process, in this case necessitating the specified ad hoc reviews.

We searched peer reviewed and gray literature from 1985-2014 over the course of these reviews. Literature was summarized for our expert panel, which met in late 2013.

Our approach to developing this measure stems from a vibrant and scientifically sound tradition regarding measuring performance. We discuss herein research involving the soundness administrative data to identify children with asthma. This is a generally accepted and standard approach with acceptable reliability.

Brook and Davies [1] trace the early history of quality measurement and remind us of the importance of medical chart audit as an approach to quality measurement. Lohr and Brook at RAND and Roos in Manitoba,

Canada pioneered the use of electronically-available administrative data (generated by routine health care operations, such as billings) as proxies for health care processes. Administrative data carefully used reduces burden of quality measurement. [2-6]

As the National Committee for Quality Assurance (NCQA) developed the Healthcare Employee Data Information Set (HEDIS) as the de facto measurement system for managed care, attention turned to the use of administrative data for routine performance measurement.

We have used rigorous and transparent methods [14] to assemble a national expert panel that included pediatricians, family physicians, pediatric and general emergency room specialists, a pediatric pulmonologist and a pediatric allergist from practices and medical schools around the country. This work was conducted in collaboration with national clinical societies (AAP, AAFP) and CAPQuaM's diverse other partner organizations, including NY State DoH/Medicaid. NCQA is an important technical consultant and partner. The specific criteria that we operationalize in this measure were all rated by the expert panel with a median score of 8 or 9 on a 9 point scale (9 high) to develop inclusion and exclusion criteria, variables for stratification and so forth. The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children.

The literature has demonstrated the reliability of claims data for assessing asthma. Though they have their limitations, these data types have been shown in multiple studies to be a reliable source of information for population level quality measurement. They are currently used for all of the analogous measures of which we are aware, including the former Core Measure and the NCQA measure considering children with persistent asthma.

The use of two years of data to validate the diagnosis of asthma has been found to produce substantial agreement with patient surveys and improves performance over the use of one year of data (28). Others have reported that using administrative databases to identify asthma is both sensitive and specific as compared to review of the primary care physician's office chart (29).

Select additional references documenting other aspects of performance gap, and supporting our process and data sources are also noted (7-13, 15-35).

1. Brook, R.H. and A. Davies-Avery, *Mechanisms for Assuring Quality of US Medical Care Services: Past, Present, and Future*. 1977: Santa Monica, CA.
2. Brook, R.H. and K.N. Lohr, *Efficacy, effectiveness, variations, and quality. Boundary-crossing research.* Med Care, 1985. **23**(5): p. 710-22.
3. Lohr, K.N., *Use of insurance claims data in measuring quality of care.* Int J Technol Assess Health Care, 1990. **6**(2): p. 263-71.
4. Roos, L.L., et al., *Using Administrative Data Banks for Research and Evaluation: A Case Study.* Evaluative Research 1979. **236**(3).
5. Roos, L.L., Jr., J.P. Nicol, and S.M. Cageorge, *Using administrative data for longitudinal research: comparisons with primary data collection.* J Chronic Dis,1987. **40**(1): p. 41-9.
6. Roos, L.L., Jr., et al., *How good are the data? Reliability of one health care data bank.* Med Care, 1982. **20**(3): p. 266-76.
7. Dresser, M.V., et al., *Clinical quality measurement. Comparing chart review and automated methodologies.* Med Care, 1997. **35**(6): p. 539-52.
8. Newton, K.M., et al., *The use of automated data to identify complications and comorbidities of diabetes: a validation study.* J Clin Epidemiol, 1999. **52**(3): p.
199-207.

9. Thompson, B.L., et al., *Measuring clinical performance: comparison and validity of telephone survey and administrative data.* Health Serv Res, 2001. **36**(4): p. 813-25.
10. Angier, H., et al., *Variation in outcomes of quality measurement by data source.* Pediatrics, 2014. **133**(6): p. e1676-82.
11. Weiskopf, N.G. and C. Weng, *Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research.* Journal of the American Medical Informatics Association, 2013. **20**(1): p. 144-151.
12. Pawlson, L.G., S.H. Scholle, and A. Powers, *Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures.* Am J Manag Care, 2007. **13**(10): p. 553-8.
13. NCQA. *National Committee for Quality Assurance*. [cited 2014 7/30/14]; Available from: http://www.ncqa.org/.
14. Brook, R.H., et al., *A method for the detailed assessment of the appropriateness of medical technologies.* International journal of technology assessment in health care, 1986. **2**(01): p. 53-63.
15. DeAngelis, C., P. Fosarelli, and A.K. Duggan, *Use of the emergency department by children enrolled in a primary care clinic.* Pediatr Emerg Care, 1985. **1**(2): p.61-5.
16. Talreja, N., et al., *Modifiable factors associated with severe asthma exacerbations in urban patients.* Ann Allergy Asthma Immunol, 2012. **109**(2): p.128-32.
17. Auger, K.A., et al., *Medical home quality and readmission risk for children hospitalized with asthma exacerbations.* Pediatrics, 2013. **131**(1): p. 64-70.
18. Ducharme, F.M., et al., *Written action plan in pediatric emergency room improves asthma prescribing, adherence, and control.* Am J Respir Crit Care Med, 2011. **183**(2): p. 195-203.
19. Farber, H.J., *Optimizing maintenance therapy in pediatric asthma.* Curr Opin Pulm Med, 2010. **16**(1): p. 25-30.
20. Smith, S.R., D.B. Wakefield, and M.M. Cloutier, *Relationship between pediatric primary provider visits and acute asthma ED visits.* Pediatr Pulmonol, 2007. **42**(11): p. 1041-7.
21. Parchman, M.L. and S. Culler, *Primary care physicians and avoidable hospitalizations.* J Fam Pract, 1994. **39**(2): p. 123-8.
22. Prevention, C.f.D.C.a. *Home-based Multi-trigger, Multi-component interventions*. 2013 [cited 2013 May 20]; Available from: http://www.cdc.gov/asthma/interventions.htm.
23. Price, J.H., et al., *Racial/ethnic disparities in chronic diseases of youths and access to health care in the United States.* Biomed Res Int, 2013. **2013**: p. 1-12; Available in open access http://dx.doi.org/10.1155/2013/787616
24. Homer, C.J., et al., *Does quality of care affect rates of hospitalization for childhood asthma?* Pediatrics, 1996. **98**(1): p. 18-23.
25. Finkelstein, J.A., et al., *Quality of care for preschool children with asthma: the role of social factors and practice setting.* Pediatrics, 1995. **95**(3): p. 389-94.
26. Flores, G., et al., *Keeping children out of hospitals: parents' and physicians' perspectives on how pediatric hospitalizations for ambulatory care-sensitive conditions can be avoided.* Pediatrics, 2003. **112**(5): p. 1021-30.
27. Knudson, A., et al., *Disparities in pediatric asthma hospitalizations.* J Public Health Manag Pract, 2009. **15**(3): p. 232-7.
28. Bindman, A.B., et al., *Preventable hospitalizations and access to health care.* Jama, 1995. **274**(4): p. 305-11.
29. National Heart, Lung and Blood Institute. *Asthma Guidelines*. 2011 February 2011 [cited 2014 7/30/2014].
30. *A Matter of Urgency: Reducing Emergency Department Overuse* in *A NEHI Research Brief* 2010, New England Healthcare Institute.
31. Martin, B.C., *Emergency medicine versus primary care: a case study of three prevalent, costly, and non-emergent diagnoses at a community teaching hospital.* J Health Care Finance, 2000. **27**(2): p. 51-65.
32. Finkelstein, J.A., et al., *Comparing asthma care for Medicaid and non-Medicaid children in a health maintenance organization.* Arch Pediatr Adolesc Med, 2000. **154**(6): p. 563-8.
33. Owens, P.L., et al., *Care of children and adolescents in U.S. hospitals*, in *HCUP Fact Book No. 4*. Agency for Healthcare Research and Quality: Rockville: MD.
34. Pearson, W.S., et al., *State-based Medicaid costs for pediatric asthma emergency department visits.* Prev Chronic Dis, 2014. **11**: p. E108.

35. Taubman, S.L., et al., *Medicaid Increases Emergency-Department Use: Evidence from Oregon's Health Insurance Experiment.* Science, 2014. **343**(6168): p. 263-268

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

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

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

_____

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

1. Evidence, Performance Gap, Priority – Importance to Measure and Report
Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***
1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form nqf_evidence_attachment_Asthma_1_12_05_16_lk.docx **1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?** Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.
1b. Performance Gap Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating: • considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or • Disparities in care across population groups. **1b.1. Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

IF a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)
IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.
Asthma is a critical problem with racial and ethnic disparities and varies by urbanicity. Adherence to the National Asthma Education and Prevention Programs (NAEPP) Guidelines improves outcomes. [1-32]. We have elsewhere provided other articles, studies, and summaries of evidence to document that ED visits and hospitalizations are typically outcome measures of choice when assessing asthma control.

ED visits for asthma in children are common and expensive. They may result from poor quality of care delivered (failure to adhere to guidelines) as well as from insufficient access to primary care. Asthma is the leading diagnosis leading to urgent care/emergent care provided in emergency departments for children. It is among the most common chronic diseases in children and expenses for asthma care are in the billions of dollars annually. Further, CMS and AHRQ assigned us this measure. In addition to data and citations provided, the team has analyzed 2007 and 2011 waves of the National Survey of Children's Health and confirmed that this parent reported measure both identified a high prevalence of asthma nationwide and significant consequences in terms of parent reported child health for children who have asthma.

Our analysis of National Survey of Children's Health [33] data (NSCH, 2011/12), estimates that 10.3 million children in the U.S. have been told that they have asthma. Of these children 7.6 million live in more urban areas that are characterized as metropolitan statistical areas (MSAs), an asthma prevalence rate of 15.4%. Table 1 shows that asthma is very consequential for health.

Table 1. Impacts of Asthma for Children Age 2-17, NSCH 2011/12
Parent/caregiver reports child's health status is excellent or very good

	2 - 5 years	6 – 11 years	12 – 17 years	Total
All Children living in Metropolitan Statistical Areas				
Asthma	59.8 %	69.6 %	74.3 %	70.1 %
No asthma	87.8 %	85.3 %	85.1 %	85.9 %
Overall	84.9 %	82.8 %	83.1 %	83.4 %
Difference	-28.0 %	-15.7 %	-10.8 %	-15.8 %
Children living in MSAs with Asthma				
All Children	59.8 %	69.6 %	74.3 %	70.1 %
Black or Latino	52.1 %	64.1 %	66.4 %	62.9 %
Not Black/Latino	66.5 %	74.6 %	80.4 %	76.1 %
Difference	-14.4 %	-10.5 %	-14.0 %	-13.2 %

We find overall a 15.8% drop in the proportion of parents who report their child's health as very good or excellent among those who have asthma, and almost twice that in younger children. Because 2 of our networks are in the greater NYC area, these data highlight children who live in more urban areas. Outside of urban areas both prevalence and gap between those with and without asthma are slightly higher (each ~17%). Effective delivery of guideline-based care can reduce the gap and decrease consequences of uncontrolled asthma, such as emergency room use and hospitalizations; better asthma care is beneficial and needed across the spectrum of children and primary care settings.[34-40] We find compelling evidence that the failure to effectively deliver guideline-based care contributes significantly to the lower health ratings for children with asthma, including for the 3.4 million urban Black and Hispanic children (age 2-17 years) with asthma. About 60% of these children are low income and have public insurance. We further are persuaded by evidence that quality of life and the quality of asthma management are associated specifically with such factors as family satisfaction with the nature of shared decision making.[41]

Citations for data demonstrating high priority

1. PCORI. PCORI Funding Annoouncement: Treatment Options for Afircan Americans and Hispanics/Latinos with Uncontrolled Asthma. 2013 [cited 2013 September 18]; Available from: http://pcori.org/assets/2013/06/PCORI-Asthma-PFA-061813.pdf.
2. Marcano-Belisario, J., Greenfield G, Huckvale K, Gunn LH, Car J, Apps for asthma self-management: a systematic assessment of content and tools. Cochrane Database Syst. Rev., 2012(8).

3. Health, O.o.M. Asthma and African Americans. [Fact Sheet]. 2012 [cited 2013 August 28]; Available from: http://minorityhealth.hhs.gov/templates/content.aspx?ID=6170.
4. Health, O.o.M. Asthma and Hispanic Americans. [Fact Sheet]. 2012 [cited 2013 August 28]; Available from: http://minorityhealth.hhs.gov/templates/content.aspx?ID=6173.
5. Wennergren, G., Strannegard I, Asthma hospitalizations continue to decrease in schoolchildren but hospitalization rates for wheezing illnesses remain high in young children. Acta Paediatr, 2002. 91(11): p. 1239-1245.
6. Wisnivesky, J., Lorenzo J, Lyn-Cook R, et al., Barriers to adherence to asthma management guidelines among inner-city primary care providers. Annals of Allergy, Asthma & Immunology, 2008. 101(3): p. 264-270.
7. DiSantostefano, R., Davis K, Yancey S, Crim C, Ecologic analysis of asthma-related events and dispensing of inhaled corticosteroid- and salmeterol-containing products. Ann Allergy Asthma Immunol, 2008. 100(6): p. 558-565.
8. Crocker, D., Kinyota S, et al., Effectiveness of home-based, multi-trigger, multicomponent interventions with an environmental focus for reducing asthma morbidity: a community guide systematic review. Am J Prev Med, 2011. 41(2): p. S5-S32.
9. Gustafson, D., Wise M, et al., The effects of combining Web-based eHealth with telephone nurse case management for pediatric asthma control: a randomized controlled trial. J Med Internet Res., 2012. 14(4): p. e101.
10. Program, T.N.A.E.a.P., Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma. J Allergy Clin Immunol, 2007. 120(5 Suppl): p. S94-138.
11. Celano, M., Holsey CN, Kobrynski LJ, Home-based family intervention for low-income children with asthma: a randomized controlled pilot study. J Fam Psychol., 2012. 26(2): p. 171-178.
12. Zuniga, G., Kirk S, et al., The impact of asthma health education for parents of children attending head start centers. J Community Health, 2012. 37(6): p. 1296-1300.
13. Blaakman, S., Tremblay PJ, Halterman JS, Fagnano M, Borrelli B, Implementation of a community-based secondhand smoke reduction intervention for caregivers of urban children with asthma: process evaluation, successes and challenges. Health Educ Res, 2013. 28(1): p. 141-152.
14. Martin, M., Catrambone CD, et al., Improving asthma self-efficacy: developing and testing a pilot community-based asthma intervention for African American adults. J Allergy Clin Immunol, 2009. 123(1): p. 153-159.
15. Seid, M., D´Amico E, Varni JW, et al., The In Vivo Adherence Intervention For at Risk Adolescents With Asthma: Report of a Randomized Pilot Trial. Journal of Pediatric Psychology, 2012. 37(4): p. 390-403.
16. Edwards, J., INSPIRE curriculum delivered in a faith-based setting. Fam Community Health., 2010. 33(2): p. 117-122.PRINCIPAL INVESTIGATOR
17. Press, V., Pappalardo AA, et al., Interventions to improve outcomes for minority adults with asthma: a systematic review. J Gen Intern Med, 2012. 27(8): p. 1001-1015.
18. DeJongh, T., Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R, Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst. Rev., 2012.
19. James, T., Fine M, Monitoring Asthma Control Using Claims Data And Patient-Reported Outcomes Measures. P.T., 2008. 33(8): p. 454-466.
20. Bender, B., Overcoming barriers to nonadherence in asthma treatment. J Allergy Clin Immunol, 2002. 109(6): p. 554-559.
21. Okelo, S., Eakin M, et al., The Pediatric Asthma Control and Communication Instrument asthma questionnaire: for use in diverse children of all ages. J Allergy Clin Immunol, 2013. 132(1): p. 55-62.
22. Benavides, S., Rodriquez JC, Maniscalco-Feichtl M, Pharmacist involvement in improving asthma outcomes in various healthcare settings: 1997 to present. Ann Pharmacother. , 2009. 43(1): p. 85-97.
23. Drotar, D., Physician behavior in the care of pediatric chronic illness: association with health outcomes and treatment adherence. J Dev Behav Pediatr., 2009. 30(33): p. 246-254.
24. Weinstein, A., The potential of asthma adherence management to enhance asthma guidelines. Ann Allergy Asthma Immunol, 2011. 106(4): p. 283-291.
25. Agency, E.P. President's Task Force on Environmental Health Risks and Safety Risks to Children:Coordinated Federal Action Plan to Reduce Racial and Ethnic Asthma Disparities. 2012 [cited 2013 August 28]; Available from: http://www.epa.gov/childrenstaskforce/actionplan_one_pager.pdf.
26. Fisher, E., Strunk RC, et al., A Randomized Controlled Evaluation of the Effect of Community Health Workers on Hospitalization for Asthma. Arch Pediatr Adolesc Med, 2009. 163(3): p. 225-232.
27. Cazzola, M., Matera M, Safety of long-acting beta2-agonists in the treatment of asthma. Ther Adv Respir, 2007. 1(1): p. 35-46.
28. Halterman, J., Fagnano M, et al., The school-based preventive asthma care trial: results of a pilot study. J Pediatr, 2012. 28(1): p. 1109-1115.
29. Association, A.L. Trends in Asthma Mortbidity and Mortality. 2012.
30. Cloutier, M., Hall C, Wakefield D, Bailit H, Use of asthma guidelines by primary care providers to reduce hospitalizations and emergency department visits in poor, minority, urban children. J Pediatr, 2005. 146(5): p. 591-597.
31. Prevention, C.f.D.C.a., Vital Signs: Asthma Prevalence, Disease Characteristics, and Self-Management Education -- United States, 2001--2009. 2011: p. 547-552.

32. Clinical Guidelines for the Diagnosis, Evaluation and Management of Adults and Children with Asthma, in New York State Department of Health and Mental Hygiene. 2008. Lawrence C. Kleinman, MD, MPH, FAAP
33. Available from: www.childhealthdata.org.
34. Diette, G.B., et al., Consistency of care with national guidelines for children with asthma in managed care. The Journal of Pediatrics, 2001. 138(1): p. 59-64.
35. Adams, R.J., et al., Impact of Inhaled Antiinflammatory Therapy on Hospitalization and Emergency Department Visits for Children With Asthma. Pediatrics, 2001. 107(4): p. 706-711.
36. Finkelstein, J.A., et al., Underuse of controller medications among medicaid-insured children with asthma. Archives of Pediatrics & Adolescent Medicine, 2002. 156(6): p. 562-567.
37. Finkelstein, J.A., et al., Self-Reported Physician Practices for Children With Asthma: Are National Guidelines Followed? Pediatrics, 2000. 106(Supplement 3): p. 886-896.
38. Bell, L.M., et al., Electronic Health Record–Based Decision Support to Improve Asthma Care: A Cluster-Randomized Trial. Pediatrics, 2010. 125(4): p. e770-e777.
39. Lob, S.H., et al., Promoting Best-Care Practices in Childhood Asthma: Quality Improvement in Community Health Centers. Pediatrics, 2011. 128(1): p. 20-28.
40. Scott, L., et al., Achieving and maintaining asthma control in inner-city children. Journal of Allergy and Clinical Immunology, 2011. 128(1): p. 56-63.
41. Gandhi, P., et al., Exploring factors influencing asthma control and asthma-specific health-related quality of life among children. Respiratory Research, 2013. 14(1): p. 1-10.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*
The overall rate of ED visits for asthma in NY State Medicaid Managed Care in 2012 is 20.65 per 100 child-years. The tables in the testing form break this down by age and race. The appendix contains additional data from the prior year including demonstrating expected seasonal variations in rate. Given our findings and our methods, although we consider this measure to be specified for a year we have demonstrated its validity to identify or compare asthma ED rate on a month-by-month basis.

The Appendix includes more data as indicated:
Page 2 Table 1. Month by Month Data, Stratified. New York State Medicaid Managed Care, 2012
Figure 2. Asthma ED Visits By Age and Month.
Page 3 Figure 3. ED Visits per 100 Child-years by Age and Urbanicity
Figure 4. ED Visits per 100 child-years by Age and County Poverty Quartile
Page 4 Figure 5. ED Visits per 100 Child Years by Age and Race/Ethnicity
Page 5 Table 3. ED Visits per 100 Child-years by Age and Quartile of Poverty
Table 4. ED Visits per 100 Child-years by Age and Urbanicity
Table 5. ED Visits per 100 Child-years by Age and Quartile of Poverty

Furthermore, within the NY State Medicaid data, a Poisson Regression Analysis and a ZIP analysis demonstrated significant differences by health plan, while controlling for Black race or Hispanic Ethnicity and Age group.

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

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*
On a yearly and a monthly basis we can demonstrate differences in the data by age, urbanicity, race/ethnicity, and level of poverty. Such differences are also evident in other cross tabulations, for example, the rate for children 2-4 in large metropolitan areas is 52.6 visits per 100 child-years compared to those in small metropolitan areas with 26.2, in micropolitan areas with 18.3

and in rural areas with 12.3. Similar magnitudes of differences were seen in other age groups, although the patterns were not all identical. Racial and ethnic differences were notable: for children ages 2-4, the rate in non-Hispanic Whites was 18.4 visits per 100 child-years, in Asians 19.3, in Hispanics 53.9 and in non-Hispanic Blacks 74. Although less dramatic, similar patterns were observed in all age groups. Overall, the rate for different races ordered by varying magnitude as illustrated between Black and White children, 41.99 and 14.79, respectively. The rate for Hispanic children was intermediate at 31.91 visits per 100 child-years. Charts and graphs are shown in the Appendix Tables 2-5 and Figures 2-5. Other disparities data has been cited elsewhere in terms of asthma control and outcomes.

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

2. Reliability and Validity—Scientific Acceptability of Measure Properties

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

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

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

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

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

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
n/a

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** Asthma_1_12_06_16-636167097129812000.xlsx

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator estimates the number of emergency department (ED) visits for asthma among children being managed for asthma. Since most hospitalizations for asthma are from the ED and many ED visits that result in hospitalization are not captured in encounter data, a numerator event may be identified either as an ED visit or as a hospitalization.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Numerator elements include the date and count of all emergency visits or hospitalizations with a primary or secondary diagnosis of asthma in a child who was eligible in the month being assessed. ED visits and hospitalizations should be identified as a visit that is associated with codes found in S.2b for identifiable asthma.

An ED visit that results in hospitalization must be counted as a single numerator event. In other words, for each individual in the denominator for the specified month, consider evidence of hospitalization that is on the same day or one day after an ED visit to represent one discrete event. Consecutive days of hospitalization are considered to represent one hospitalization.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator represents the person time experience among eligible children with identifiable asthma. Assessment of eligibility is determined for each child monthly. The total number of child months experienced is summed and divided by 1200 to achieve the units of 100 child years for the denominator.

Assessing eligibility for the denominator requires 2 years of data, the reporting year and the 12 month period before the reporting year. (See Appendix 1, Figure 1)

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The denominator is the sum total of the number of months that children meet all eligibility criteria divided by 1200. This calculation yields the denominator in terms of '100 child years', which is the equivalent of 100 children with identifiable asthma in the plan for 1 year each.

We consider children to be managed for identifiable asthma to meet two criteria simultaneously:
1) They have been enrolled for three consecutive months including the month being assessed, and
2) They have evidence of claims sufficient to meet the eligibility criteria for identifiable asthma.

The analysis should be conducted on a month by month basis as described herein: Within the group of children who meet the criteria for identifiable asthma, identify and maintain a unique patient identifier, age, and all stratification variables. We call the time frame during which eligibility is established to be the Assessment Period.

For each month of the Reporting Year, determine eligibility for each patient, as of the last day of the month prior to the reporting month. This illustration assumes that the Reporting Year is 2011. When assessing January 2011, consider all of Calendar Year 2010 as the Assessment Period for assessing the presence or absences of identifiable asthma. For February, 2011 the Assessment Period includes all of calendar year 2010 AND January 2011. Repeat this progression monthly so that for December, 2011 identifiable asthma one would identify children with identifiable asthma using an Assessment Period from January 2010 through November 2011. For each month, assess whether the continuous enrollment criterion is met prior to including the month in the denominator. For example, for January 2011, the child must have been enrolled in November and December, 2010 (plus January 2011). Another example, for December 2011, to be eligible the child must have been enrolled in October 2011 and November 2011, as well as December.

Please see Appendix: Figure 1 and codes used for definitions (Sb.2). These are considered INTEGRAL to these specifications and are not optional.

Identifiable asthma is present when there is evidence as specified for any of the following:
a. Prior hospitalization with asthma as primary or secondary diagnosis
b. Other qualifying events after the fifth birthday (age is age at occurrence):
i. One or more prior ambulatory visits with asthma as the primary diagnosis AND a subsequent ED visit in the Reporting Month, OR
ii. Two or more ambulatory visits with asthma as a diagnosis, OR
iii. One ambulatory visit with asthma as a diagnosis AND at least one asthma-related prescription, OR
iv. Two or more ambulatory visits with a diagnosis of bronchitis
c. Other qualifying events, any age:
i. Three or more ambulatory visits with diagnosis of asthma and/or bronchitis, OR
ii. Two or more ambulatory visits with a diagnosis of asthma and/or bronchitis AND one or more asthma- related prescriptions.

As noted in the specifications, asthma-related medicine means long-acting beta-agonist (alone or in combination) or inhaled corticosteroid (alone or in combination), anti-asthmatic combinations, methylxanthines (alone or in combination), and/or mast cell stabilizers.

Please note that in order to promote better harmonization, we start with the current HEDIS asthma medication list. From that list, in accordance with our expert panel recommendations we eliminate medications in the following 2 categories: leukotriene modifiers, short-acting inhaled beta-agonists.

We further exclude indacaterol, a recently approved long acting beta agonist that is indicated in the US only for the treatment of COPD. As indicated elsewhere, COPD is an exclusion criterion for this measure. These specifications anticipate that NCQA will update the medication list from time to time and with the stated exclusions updated lists may be substituted for the list linked herein. The table used for testing is labeled Table AMR-A: Asthma Controller and Reliever Medications, and can be found at http://www.ncqa.org/HEDISQualityMeasurement/HEDISMeasures/HEDIS2015/HEDIS2015NDCLicense/HEDIS2015FinalNDCLists.aspx (last accessed September 12, 2015).

If pharmacy data are not available, the measure should be reported with notation that pharmacy data were not used for the assessment of eligibility. This avoids eliminating from the measure those facilities with no link to pharmacies. Our testing reveals that only a small proportion of patients are excluded by not including pharmacy data to establish eligibility.

The presence of identifiable asthma (see S.2b and above) is established each month from administrative data using the specified algorithm. (Appendix: Figure 1 and this section's narrative)

All events in the administrative data should be associated with a date of service.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*
Children with specified concurrent or pre-existing diagnosis and children who have not been consecutively enrolled in the reporting plan for at least three months, as specified in the details section.

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
Children with concurrent or pre-existing: Chronic Obstructive Pulmonary Disease (COPD) diagnosis, Cystic Fibrosis diagnosis, or Emphysema diagnosis.

Children who have not been consecutively enrolled in the reporting plan for at least two months prior to the index reporting month and for the reporting month (a total of three consecutive months ending in the reporting month).

For entities that use AHRQ's Clinical Classifications Software, please note that it is important to apply the exclusion after identifying visits that satisfy CCS class 128.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and*

coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)
This measure calls for stratification by age group, by race/ethnicity, and by age group and race/ethnicity. Several additional stratifications are recommended but optional. These may be required by the accountability entity or reported by the reporting entity. These variables include rurality/urbanicity and county level of poverty.

Age groups are 2-5, 6-11, 12-18, and 19-20, each inclusive. (reporting entity should specify whether to use age at month of qualifying event or age on first day of reporting year)

Race/ethnicity should incude White non-Hispanic, Black non-Hispanic, and Hispanic as well as other groups as requested by the accountability entity and consistent with current HHS usage.

For social demographic stratification: identify County equivalent of child's residence. If County and State or FIPS code are not in the administrative data, the zip codes can be linked to County indirectly, using the Missouri Census Data Center (http://mcdc.missouri.edu/). These data will link to County or County equivalents as used in various states.

i.Identify the Urban Influence Code (1) or UIC for the county of child's residence. (2013 urban influence codes available at: http://www.ers.usda.gov/data-products/urban-influence- codes.aspx#.UZUvG2cVoj8).

ii.Identify the Level of Poverty in the child's county of residence. The percent of all residents in poverty by county or county equivalent are available from the US Department of Agriculture at http://www.ers.usda.gov/data-products/county-level-data-sets/download- data.aspx. Our stratification standards are based on 2011 US population data that we have analyzed with SAS 9.3. Using child's state and county of residence (or equivalent) or FIPS code, use the variable PCTPOVALL_2011 to categorize into one of 5 Strata:

a.Lowest Quartile of Poverty if percent in poverty is <=12.5%
b.Second Quartile of Poverty if percent in poverty is >12.5% and <=16.5%
c.Third Quartile of poverty if percent in poverty is >16.5% and <=20.7%
d.First Upper Quartile (75th-90th) if percent in poverty is >20.7% and <=25.7%
e.Second Upper Quartile (>90th percentile)

iii.Categorize age by age at the last day of the month that ends the assessment period. Aggregate into age categories 2-4, ages 5 through 11, ages 12-18, ages 19-21.

iv.Categorize Race/Ethnicity as Hispanic, Non-Hispanic White, Non- Hispanic Black, Non-Hispanic Asian/Pacific Islander, and Non-Hispanic Other

v.Categorize Insurance Type as Private (Commercial), Public, None or Other

vi.Categorize benefit type as HMO, PPO, FFS, PCCM, or Other

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
Other
If other: In order to allow for more granular comparisons this measure is specified to be stratified. Stratification for risk adjustment of this measure would not be justified by the literature. Although epidemiological findings support our stratification schema, no biological evidence exists to support intrinsic correlation of ED rates with stratification variables.

S.12. **Type of score:**
Rate/proportion
If other:

S.13. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Lower score

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

Step 1: Measure person-time eligible for each patient and record by month.
a. For each month in the reporting year, identify all children ages 2 – 21 years who meet the criteria for Identifiable asthma during the assessment period. The assessment period is defined as the year prior to the reporting year plus all months in the reporting year prior to the reporting month.

Identify and maintain a unique patient identifier and all stratification variables.

To illustrate: if the goal is to report for January 2011, first one would identify children with Identifiable asthma using the criteria, and analyze all of calendar year 2010 when doing so. Continuous enrollment criterion requires that the child was enrolled in November and December of 2010, as well as January 2011. This total represents the number of person-months (child-months) for January.

Next, for February: one would identify children with Identifiable asthma using the criteria, and analyze all of calendar year 2010 AND January 2011 when doing so. Continuous enrollment criterion requires that the child was enrolled in December 2010 and January 2011, as well as February 2011. This is the number of person-months (child-months) for February.

Repeat this progression monthly so that for December, one would identify children with Identifiable asthma and analyze all of calendar year 2010 AND January through November 2011 when doing so. Continuous enrollment criterion requires that the child was enrolled in October 2011 and November 2011, as well as December 2011. This is the number of person-months (child-months) for December.

b. Sum all months that are eligible from the reporting year. This sum is the denominator in people-months. Divide by 1200. This is denominator in 100 people-years. This is the denominator for the year.

Step 2: Month by month, considering the definitions above, identify the number of discrete numerator events that occur in children eligible in that specific month:
a.Identify the number and date of ED visits with asthma as a primary or secondary diagnosis among those children who are eligible for that reporting month.
b.Identify the number and date of inpatient hospitalizations with asthma as a primary or secondary diagnosis among those children who are eligible for that reporting month.
c.Identify the number of discrete numerator events. Consecutive days with inpatient hospital codes are considered one hospitalization. Hospitalizations on day of or day after ED visit are NOT considered discrete from the ED visit.
d.Sum the number of numerator events across the year.
e.Maintain stratification variables and unique identifiers.

Step 3. Calculate rate as Numerator / Denominator. While this measure is specified for the year, it has also been validated to demonstrate seasonality using monthly rates.

Step 4. Calculate stratification variables as specified in S.12.

Step 5. Repeat by strata. Within age strata repeat by other specified strata. Perform other cross tabulations as requested by the accountability entity. Eliminate any strata with less than 40 person-months in any month's denominator OR less than 1000 person-months for the year.

Appendix 1: Figure 1 illustrates the calculation of person-time and is considered fundamental to this calculation algorithm.

When data cannot be obtained from any source:
- If critical for calculation – delete patient from consideration for that reporting month
- If non-critical for calculation – include patient
Critical data include encounter data for the reporting month and some period of time in the assessment period. In order to report stratifications age and race/ethnicity are considered critical. Pharmacy data are not considered critical

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
n/a

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
n/a

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Claims (Only), Claims (Other)

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
n/a

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
No data collection instrument provided

S.20. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Health Plan, Population : Community, County or City, Population : Regional and State

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Emergency Department, Hospital, Hospital : Acute Care Facility, Other
If other: This measure incorporates data from the ambulatory, ED, and hospital settings to describe performance at the level of the plan or the community.

S.22. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
n/a

2. Validity – See attached Measure Testing Submission Form
nqf_testing_attachment_Asthma_1_12_07_16_lk_v2.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.3 For maintenance of endorsement
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (e.g., potential factors identified in the

literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care)

What were the statistical results of the analyses used to select risk factors?

Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)
No - This measure is not risk-adjusted

**Note: New information added since the previous review is in blue text

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Rate of Emergency Department Visit Use for Children Managed for Identifiable Asthma: Visits per 100 Child-years

Date of Submission: 12/14/2016

Type of Measure:

☒ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process	☐ Efficiency
☐ Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure*. *Testing must be provided for all the sources of data specified and intended for*

measure implementation. ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.23***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ administrative claims	☒ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

New York State Medicaid claims data.

Also, our work builds off of work performed by our CAPQuaM partner and steering committee member, NCQA. For specific data reliability and signal to noise analyses, we incorporate by reference (and will present more selectively) NCQA data relevant to their submission for NQF –endorsed asthma related measures:

- Use of Appropriate Medications for People with Asthma (ASM) – 0036 (we understand this is no longer being maintained as of 2015, but it was endorsed and the data were accepted.)
- Medication Management for People With Asthma (MMA) – 1799
- Asthma Medication Ratio (AMR) – 1800

We note that 1799 and 1800 are not directly applicable because they were tested at the score level. However, the scores were dependent upon definitions which use the same data element level as our measure and thus provide indirect evidence of the capacity of a measure using such data elements to produce valid scores.

The analyses above provide information regarding the capacity to use administrative data to identify the applicable denominator population. There is nearly complete overlap of the denominator codes and there is overlap of the denominator elements. Where codes differ it is specific to decisions made by the CAPQuaM expert panel which was aware of the NCQA measures.

1.3. What are the dates of the data used in testing? 2010 - 2012

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (***must be consistent with levels entered in item S.26***)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☒ health plan	☒ health plan
☒ other: Population, State, Region, County, Integrated delivery system	☒ other: Population, State, Region, County, Integrated delivery system

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

NY State Medicaid Managed Care claims data, including claims from all MCO's that are contracted for Medicaid care by our partner, the NY State Department of Health.

For our primary analysis of MCO's we included both full (8) and partially (10) capitated plans, each of which had at least 900 children who contributed time to the denominator of the measure.

For our primary analysis of county of residence, we used 45 counties that contributed at least 1000 months of person time to the denominator.

The numbers we present are from reporting year 2012, include children from counties in nine urban influence codes and in counties poverty level 1-3. NY State does not have any counties in the lowest 25% of poverty or with UIC of 10-12. New York has more than 60 counties and numerous health plan vendors. Analysis in Year 2011 provided very similar data.

Foundational analyses for this measure were performed and previously reported by NCQA considering *nine health plans covering a variety of geographic areas within the United States that were asked to provide a complete administrative data file consisting of any member in their commercial and Medicaid product lines for anyone that had a diagnosis code for asthma during the calendar years of 2009-2010. The complete member-level administrative file used for analysis included a total of more than 82,000 health plan members with asthma*.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

All children 0-21 with records in the 2011 in the 2011 NYS Medicaid Managed Care (MMC) administrative database and all pediatric patients meeting the criteria for identifiable asthma in the 2012 NYS Medicaid Managed Care administrative database.

There were 192,722 children with identifiable asthma in the managed care (plan level) analysis, 211,703 in the county analysis, and 212,432 overall in MMC.

19,903 children experienced 30,382 qualifying emergency department visits in the reporting year. 1806 visits were in young adults age 19 or 20. In 2011, the median number of visits per child with an ED visit was 1, the 75th percentile was 2, and the 90th percentile was 3. One percent had 6 or more visits.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Data source 1 (Chart):
Prior to initial specification of the measure we contracted for a survey of quality managers representing more than a dozen hospitals to assess data availability and the ease and feasibility of abstraction of data relevant for asthma and other CAPQuaM measure, including 10 hospitals that responded regarding asthma-specific data elements. Our survey found that availability in the medical record of age, (date of birth), race, ethnicity, date and site of visit, documentation of primary or secondary diagnosis of asthma, hospitalization, and payment source were routinely available in the chart and "Not Difficult to Collect." Our chart review of 1200 medical records for ED visits in a single institution performed for validation of a sister measure on appropriateness confirmed the availability of these data. This validates the capacity to obtain such data form the medical record and the primary occurrence of the data in the chart so that coders have the clinical information required to population ICD-9/10, CPT, and Revenue codes that comprise the administrative data that are the preferred data source for these measures.

Data Source 2 (Administrative):
Assessment of the capacity to identify the eligible population and qualifying events was performed in NY State Medicaid data in both 2011 and 2012 reporting years.

For MCO analysis we analyzed both with the 18 plans that had 900 or more children contributing to the denominator and with the 20 plans that contributed at least 1000 months of person time to the denominator and found no meaningful differences in the analyses. We present details from the analysis of the 18 plans.

For county level analysis we included those 45 counties that contributed at least 1000 months of person time to the denominator.

Other analyses included all children with identifiable asthma.

Data source 3 (explicit criteria):
Our construct for the CAPQuaM measure was defined by the multidisciplinary national expert panel using a RAND type modified Delphi process, which produced a set of explicit criteria that were both substantive and addressed specification details, such as what combination of administrative codes could be used to identify a child with asthma, other inclusion and exclusion criteria for the measure, and preferences regarding how to report and stratify the measure.

The panel initially used the term persistent asthma to describe asthma that was pre-existing and should have been recognized as asthma by the health care system prior to the timing of the ED visit. This construct was renamed by our stakeholder group to be identifiable asthma to avoid confusion with other uses of the term persistent asthma. The construct was intended to be more inclusive than HEDIS' persistent asthma diagnosis, while still removing from consideration those whose asthma was unlikely to have been actively managed at the time.

Data Source 4 (National Survey of Children's Health)
We validate the construct of identifiable asthma comparing it to two other constructs:
HEDIS' definition of persistent asthma, which should have been more restrictive than 'identifiable asthma'; and the National Survey of Children's Health's question regarding if the caregiver had ever been told by a doctor or nurse that the child had asthma, which should have been less restrictive than 'identifiable asthma.' The former analysis was conducted in Medicaid 2011 and the latter in the most recent NSCH data.

Holding steady the continuous enrollment criterion at 12 months, HEDIS criteria identified a rate of persistent asthma of 3.1% with the CAPQuaM criteria identifying identifiable asthma at a rate of 8.6%. As expected, identifiable asthma was between 2 and 3 times more permissive than the intentionally restrictive persistent asthma. We analyzed NSCH data to estimate a population rate of asthma in NY State Medicaid child population to be between 15 - 16%, indicating that our criteria did provide a meaningful filter as we had intended.

Reducing the continuous enrollment period down to three months as was suggested by members of our steering committee increases the number of children eligible for the measure by several tens of thousands while still restricting the measure to those who had received sufficient care for asthma to be identified, and requiring continuous enrollment for attribution to the extent felt important by our multi-stakeholder group. This inclusiveness help to counter risks of churning that are particularly prominent in the Medicaid population. This analysis was conducted in the NYS Medicaid data.

Data Source 5: HEDIS
Assessment of data elements for identifying a population with asthma and asthma scores was performed by NCQA in nine geographically diverse managed care plans. We considered the HEDIS data for measures 1799, 1800 and 0036. We cite 1799 and 1800 not as specific evidence of score level performance of our measure, but as evidence that measures that rely on the same administrative data elements for their denominator have the capacity to distinguish signal from noise at a very high level.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Race, ethnicity, zip code/county of residence, level of poverty in the county of caregiver residence, and urban influence code for the county of caregiver residence for the NY State analysis. Within the Medicaid data, we looked at eligibility category.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

See section 2b2 for validity testing of data elements.

See 2b5 for testing that demonstrates sufficient precision and reliability of the performance measure score. Please note as this is a rate the binomial signal to noise ratio analysis is not appropriate for this measure. Similarly since measured variation between entities is expected sometimes to be meaningful and other times not, the key performance attribute is to demonstrate the capacity in real use to identify which of multiple comparisons are statistically significant. We discuss our findings in this context in 2b5.2.

In addition to the analyses presented below, we also conducted the analysis using a zero inflated poisson approach.

With NY State Medicaid we conducted analyses that demonstrate the measure's capacity to distinguish among health plans. The standard approach to measuring reliability is inappropriate as the measure is a rate and not a binomial. The appropriate model is either a Poisson, model (which is discussed in 2b5), a hurdle model or a Zero inflated Poisson (ZIP). Hurdle requires additional assumptions that model two processes, and is more sensitive. ZIP misses out on capturing some of the plans' impact on whether a child makes it to the ER, but models the rate very well. We performed both with similar results and report on the ZIP as the more conservative approach (it under attributes the impact of the plan).

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

ZIP Models: Using Proc HPFMM with a log link, a Poisson distribution and an offset equal to the log of the number of months the child had asthma in the plan, the model was highly significant (p<.0001) incorporating specified age groups and plans as categorical variables. Comparing to a randomly selected index plan, 14 of 17 plans had statistically significant differences in performance with the median and modal p-value being <0.001. Non-significant plans' p-values=0.08, 0.16 and 0.88.The model is able to differentiate distinct performance levels. Results were similar when we performed the models considering only plans, after stratifying for age group. Because of low numbers in the 18-21 yr old group across plans, fewer were significant, but findings suggest that the measure is sensitive to real differences given adequate sample sizes.

Ages 2-4: 15 plans of 17 are significant (p<0.05). Additional are 0.06 and 0.21.
Ages 5-11: 14 plans of 17 are significant (p<0.05). Additional are 0.37, 0.21, and 0.70.
Ages 12-18: 13 plans of 17 are significant (p<0.05). Additional are 0.11, 0.06, 0.26, and 0.43.
Ages 19-21: 7 that were significant (p<0.01). In general the sample size was sufficient to assess some plan's performance for this group.

ZIP models also showed that even after controlling for age groups: Urban counties have different performance than rural counties; Large urban counties are distinct in performance from all others; Small urban counties are different from suburban counties and rural counties, although the smaller numbers in rural counties contributes to a P-value of 0.07; Performance in suburban and rural counties are generally similar. New York State does not have extremely rural counties; ED utilization of Blacks is significantly different from Whites (p<0.01); ED utilization of Hispanics is significantly different from Whites (p<0.01); ED utilization of Blacks and Hispanics are significantly different from one another (p<0.01).
These data contribute evidence to support use of the measure, adding both to the data on reliability (as plan to plan differences were meaningful) and validity (in that the models performed as predicted and consistent with current knowledge regarding variations associated with race, ethnicity, and urbanicity).

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The ZIP models reinforce what is described below and add robustness to our interpretation that the performance measure scores demonstrate reliability with a high degree of certainty and confidence.

_____

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☐ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Analysis of 2011 and 2012 data provided similar findings, including the pattern of month to month variations and the variability associated with various stratification variables, demonstrating test-retest reliability. Further the identification of predicted seasonal changes within a defined population is a more difficult challenge for distinction (or signal to noise) than comparison in distinct populations and confirms in the analysis theoretically sound and predicted differences among groups of children. We found evidence of theoretically sound and predicted differences whether we categorized on race (Blacks were highest rate), type of health plan (fully capitated pans had lower rates than partially capitated plans), poverty (higher poverty had higher rates), or urbanicity (large urban areas had the highest rates), and age group (school age had the lowest rates). This leads us to have high confidence that the performance measure scores are a valid and reliable indicator of quality and of the underlying construct of undesirable outcomes in asthma.

Observed differences between counties and between health plans were similar regardless of whether or not we controlled for age group, race, ethnicity in the analysis.

Please see descriptions of both NCQA and CAPQuaM testing above in 1.2-1.7.

The literature also supports the use of claims data to identify the presence of asthma. We use administrative data to identify the age of the child, various stratification variables and the presence of asthma, as well as the presence of an asthma ED visit or hospitalization. These are routinely used to support billing by CMS, Medicaid, and private insurers and are routinely used in quality measurement.

There is moderate agreement (kappa = 0.45 – 0.50) when comparing administrative data regarding the presence of constructs such as recent asthma attacks, use of asthma medications, attack or medication, attack and medication, using 1 year of administrative claims data. The agreement improves from 0.55 to 0.60 when using two years of data as this measure does. (8). We expect that these kappas would be significantly higher were the analyses restricted to children with disease that met our construct criteria for identifiable asthma.

We followed a peer-reviewed systematic process for measure development that incorporated a literature review, an expert panel process, and a multi-stakeholder process that included the input and review of the measure by potential users in the development process. More details about this process are available upon request.

Explicit criteria were developed using a variation of the two-round modified Delphi process RAND/UCLA Appropriateness Method with a multidisciplinary and geographically diverse expert panel comprised of both clinicians and researchers. Identifiable asthma was based on panel findings and appropriateness criteria included for this measure were those that were both available in the chart and highly rated. The general reliability of this approach is well established. [9, 10] It has been applied successfully to pediatric services previously. [11-13] The expert panel further validated the measure subsequent to development via an email poll.

Development included a series of alpha tests to refine specifications by conducting iterative analyses in New York State Medicaid data. Conclusions from alpha tests include:

1) The reporting period and the assessment period could not overlap completely, leading to use of 2 years of data as shown in the specifications' diagram. The optimal approach was to divide the reporting year into 12 reporting months. ED events in that month are eligible for the numerator if persistent asthma criteria have been satisfied (combining the look-back year and all prior months in the reporting year) and the child has been continuously enrolled for the two months immediately prior to the reporting month. The optimal building block unit for the denominator is in child-months, which is rolled up to child-years;

2) Using both revenue codes and CPT codes increased our sensitivity meaningfully, a choice validated by consultation with coding and billing experts and confirmed by analyzing the NY State data;

3) NY State Medicaid data and national survey data (HCUP) converged to demonstrate the importance of including hospitalizations as numerator events even when the underlying construct is ED visits. This is consistent with policies of many payers to request providers not to submit both ED and hospital claims for the same day. Error is far less by considering both ED visits and hospitalizations as numerator events, than by not including hospitalizations. (See poster that follows in the next section)

4) The expert panel only wanted numerator events for which the children were already known to the accountable entity as having asthma and established definitions for such "identifiable asthma". Alpha testing in NY State Medicaid demonstrated the expected results:

    a. Holding steady the continuous enrollment criterion at 12 months, HEDIS criteria identified a rate of persistent asthma of 3.1%, the CAPQuaM criteria identifying identifiable asthma at a rate of 8.6%. This more inclusive approach was our goal.

b. More than 25% of children with any asthma claim are not included in the denominator, indicating that this is a meaningful filter. Confirming this, the observed rate of 8.3% in the denominator. 8.3 is just over half of what we found when analyzing NSCH data to identify an expected rate of NY State Medicaid children whose caregivers would report that they every been told the child had asthma.

c. Relaxing the continuous enrollment period to 3 months was suggested by members of our stakeholder steering committee. Doing so increased the eligible number by more than 20,000 while still restricting the measure to those who had received sufficient care for asthma to be identified, and requiring continuous enrollment for attribution to the extent felt important by our multi-stakeholder group.

The use of Expert Panels has been demonstrated to be useful in measure development and health care evaluation, including for children. [14]

The definitions were specified to allow their use with data elements that ought to be available in electronic form to a responsible entity, such as a health plan or state Medicaid program. Potential exceptions to this are elements such as ZIP code of residence and race and ethnicity of the child. We have data from a feasibility study we conducted with a contractor that surveyed quality departments at more than a dozen hospitals across three measure sets. 10 hospitals responded to the asthma-specific questionnaire. We found that these data elements are generally available in the chart, although the definition of race and ethnicity, as well as how it is determined, may vary by institution. Nonetheless, the CHIPRA legislation (2009), which has funded the development of this measure, directs for measures to be capable of identifying disparities and we have specified it to be so, despite concerns about reliability in the collection and assessment of race and ethnicity by health-care-providing institutions and practices. In this case, we need to drive performance through measurement, as it is foundational to the legislative and executive branch sources of our funding.

former Medicaid core measure that we were tasked with enhancing was a simple risk, with asthma patients defined in the measurement year as having primary or secondary diagnosis for any service, and ED visits defined as CPT-code-identified ED visits with asthma as the primary diagnosis. The numerator for the Core Measure includes all patients with at least one ED visit for asthma as asthmatic events, whether or not the patient was known to be an asthmatic before the event. Further, numerator events alone could qualify children for inclusion in the denominator. Our partners in the New York State Medicaid program have described this characteristic as highly undesirable and the CAPQuaM team agreed, prompting our month–by-month approach to analysis. We enhanced the validity of this measure by deflating competing concepts and clearly specifying it as an interpretable epidemiological rate (incidence density). Enhancements include: we set a threshold of utilization below which a child is not considered to have given the health care system an opportunity to have identified the child as an asthmatic; we restrict the measure to those children who meet this threshold before the ED visit occurs and we are measuring and incidence density or rate and not a risk, allowing us to count each ED visit in the numerator and person-time in the denominator. While the median number of visits among those with visits is 1, more than one-quarter of children in New York State Medicaid Managed Care with an ED visit have a second visit. A few outliers contribute more than 10 Ed visits per child. The rate measure allows us to provide a better estimate of the number of undesirable outcomes, rather than the number of children with undesirable outcomes.

As a rate, one child can contribute to the numerator many times. It also is self-adjusting for children who enter or leave the eligible population since children contribute to the denominator independently for each month that they are eligible. It also assures that ages can be calculated to the month rather than to the year, if the reporting entity requests this level of detail. Our analysis uses the age on the first day of the Reporting Year.

To enhance the meaningfulness of the measure, we have included a two-month continuous enrollment requirement prior to the reporting month. Since the child must also be eligible for the reporting month, this becomes a three-month continuous enrollment requirement. In doing this, we sought to strike a delicate balance between developing a meaningful accountability measure and eliminating children because of problems of churning, which have been well

documented by researchers (15). This balance was achieved in close collaboration with our colleagues at NY State Medicaid.

The development team's goal was to develop an ICD10 code set that was fully consistent with the intent of the original measure. Our process began by performing general equivalency mapping using the forward mapping from www.icd9data.com. We then did a de novo review of the CMS ICD 10 CM set to seek to identify codes that might be appropriate for asthma. We reviewed potential codes identified by both sources and developed a new list of codes appropriate for inclusion criteria and a new list of codes appropriate for exclusion criteria. Drs. Kleinman and Sharma reviewed the lists independently and then achieved consensus in a conference call review and discussion. The guidance for the intended constructs for both ICD9 and ICD10 coding were the findings from a RAND style modified Delphi panel that incorporated 9 national experts over the course of the measure development process.

1. Dresser, M.V., et al., *Clinical quality measurement. Comparing chart review and automated methodologies.* Med Care, 1997. **35**(6): p. 539-52.

2. Newton, K.M., et al., *The use of automated data to identify complications and comorbidities of diabetes: a validation study.* J Clin Epidemiol, 1999. **52**(3): p.

   199-207.

3. Thompson, B.L., et al., *Measuring clinical performance: comparison and validity of telephone survey and administrative data.* Health Serv Res, 2001. **36**(4): p.

   813-25.

4. Angier, H., et al., *Variation in outcomes of quality measurement by data source.*

   Pediatrics, 2014. **133**(6): p. e1676-82.

5. Weiskopf, N.G. and C. Weng, *Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research.* Journal of the American Medical Informatics Association, 2013. **20**(1): p. 144-151.

6. Pawlson, L.G., S.H. Scholle, and A. Powers, *Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures.* Am J Manag Care, 2007. **13**(10): p. 553-8.

7. NCQA. *National Committee for Quality Assurance*. [cited 2014 7/30/14]; Available from: http://www.ncqa.org/

8. Huzel, L, et al. Diagnosing Asthma: The fit between survey and administrative database. Can Respir J. 2002 Nov-Dec;9(6):407-12.

9. Fitch, K., et al., *The RAND/UCLA Appropriateness Method User's Manual*. 2001 RAND.

10. Kosecoff, J., et al., *The appropriateness of using a medical procedure. Is information in the medical record valid?* Med Care, 1987. **25**(3): p. 196-201.

11. Kleinman, L.C., et al., *The medical appropriateness of tympanostomy tubes proposed for children younger than 16 years in the United States.* Jama, 1994. **271**(16): p. 1250-5.

12. Kleinman, L.C., E.A. Boyd, and J.C. Heritage, *Adherence to prescribed explicit criteria during utilization review. An analysis of communications between attending and reviewing physicians.* Jama, 1997. **278**(6): p. 497-501.

13. Keyhani, S., et al., *Overuse of tympanostomy tubes in New York metropolitan area: evidence from five hospital cohort.* Bmj, 2008. **337**: p. a1607.

14. Brook, R.H., et al., *A method for the detailed assessment of the appropriateness of medical technologies.* International journal of technology assessment in health care, 1986. **2**(01): p. 53-63.

15. Fairbrother, G., et al., Churning in Medicaid managed care and its effect on accountability. J Health Care Poor Underserved, 2004. 15(1): p. 30-41.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Performance Measure Score:

We found statistically meaningful differences in predicted directions when we used categories such as race/ethnicity, age group, month of year, type of managed care, level of poverty, and urbanicity. These differences were robust to whether we considered them individually or in a common model. The model for testing as a Poisson regression with a log-link function in SAS (Proc GenMod), using the number of ED visits as the outcome and the natural log of the months of exposure as the offset.

When we tested for plan to plan differences findings were similar. If we randomly selected a plan we typically found more than half differing from it with a P<0.05, typically less than 0.01. If we picked extreme plans, virtually all were different. We found similar findings when using county rather than plan. (See section 2b5.2). Statistical differences were robust to inclusion of the age group and race/ethnicity. Stratified by age group, similar findings were found, with somewhat fewer plan-plan and county to county differences found in the older age group, which had smaller numbers of children.

Data Elements and Expert Process

For the foundational NCQA work (Measures 1799, 1800, 0036), NCQA's field test retested a number of previously validated criteria for identifying an eligible population with persistent asthma using administrative claims data. Using the dataset provided, NCQA examined several different scenarios to determine the effects of different specification criteria on this particular population. This information was combined with multiple years of HEDIS data collection of this measure to examine the reliability of collecting this measure through administrative claims.

From NCQA's submissions: *Reliability was estimated on the HEDIS 2011 submissions (2010 data) using the beta-binomial model. Beta-binomial is a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® health plan measures. The beta-binomial model assumes the plan score is a binomial random variable conditional on the plan´s true value that comes from the beta distribution. The beta distribution is usually defined by two parameters, alpha and beta. Alpha and beta can be thought of as intermediate calculations to get to the needed variance estimates. The beta distribution can be symmetric, skewed or even U-shaped.*

Reliability used here is the ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

For the foundational NCQA work, NCQA's field test retested a number of previously validated criteria for identifying an eligible population with persistent asthma using administrative claims data. Using the dataset provided, NCQA examined several different scenarios to determine the effects of different specification criteria on this particular population. This information was combined with multiple years of HEDIS data collection of this measure to examine the reliability of collecting this measure through administrative claims. They report that score level reliability of the HEDIS 2011 submissions (2010 data) was assessed using the beta-binomial model. Beta-binomial is a better fit when estimating the reliability of simple pass/fail rate measures as is the case with most HEDIS® health plan measures.

For the NCQA analysis reliability was reported as the ratio of signal to noise. The higher the reliability score, the greater is the confidence with which one can distinguish the performance of one plan from another. A reliability score greater than or equal to 0.7 is considered very good.

We note that 1799 and 1800 are not directly applicable because they were tested at the score level. However, the scores were dependent upon definitions which use the same data element level as our measure and thus provide indirect evidence of the capacity of a measure using such data elements to produce valid scores.

Thus we cite them not as specific evidence of our score level performance of the submitted measure, but as evidence that the HEDIS measures that rely on the same administrative data elements for their denominator have the capacity to distinguish signal to noise at a very high level. While the evidence is indirect it is dispositive. That is, we assert that had the data elements been inadequate it would result in non-differential misclassification error which is a major bias towards the null thus introducing noise and reducing signal. That this does not happen to an appreciable degree specifically implies that the data elements function well – indeed this could be one rationale for why NQF allows the use of performance score level analysis in the first place. These findings provide strong indirect evidence of the validity of our approach to capturing the measure's denominator.

There is nearly complete overlap of the denominator codes and there is overlap of the denominator elements. Where codes differ it is specific to decisions made by the CAPQuaM expert panel which was aware of the NCQA measures. Review of the medication lists for 0036 reveal that all medication used by the submitted CAPQuaM measure are also in the HEDIS measure. The CAPQuaM measure excludes specifically short acting beta agonists and leukotriene inhibitors at the specific direction of the CAPQuaM expert panel. We also specify exclude indacaterol from the list of "asthma specific medications" since it is a long acting beta agonist which is only indicated in the USA for treatment of COPD, which is a specific exclusion criterion for this measure.

Further, we identify asthma visits and medications using the same data that an insurance company or Medicaid would use for payment, including ICD codes, CPT codes, and revenue codes. We have had conversations with expert coders and New York State Department of Health Office of Health Insurance Programs to confirm our choices. Our literature review found that while there is moderate agreement (kappa = 0.45 - 0.50) when comparing administrative data regarding the presence of constructs such as recent asthma attacks, use of asthma medications, attack or medication, attack and medication, using 1 year of administrative claims data to parent report, the agreement improves from 0.55 to 0.60 when using two years of data.(1) We expect that these kappas would be significantly higher were the analyses restricted to children with disease that met our construct criteria for identifiable asthma.

The literature further supports our work. ICD-9 and ICD-10 codes for asthma on patients' medical charts typically match claims data. ICD-9-CM administrative data have been validated using various methodologies for various purposes (2-10). As examples: Jollis et. al. compared insurance claims data to the clinical database data to identify patients using ICD-9-CM codes for selected diagnoses and found that when all diagnoses were included, overall kappa agreement was .75 (2). Lee et. al. compared heart failure diagnoses identified in ICD-9 to the Framingham clinical criteria as the gold standard and found a positive predictive value of 94.3% (3). Muhajarine et. al. compared self-reported heart health survey data to physician claims from a database registry and found an overall agreement for hypertension of 81.7% indicating moderate to high agreement(4).Quan et. al. tested administrative discharge data to chart data for recording of comorbidity information using a Charlson index for measurement. Overall agreement of the Charlson index was good between databases but decreased as burden of comorbidity increased. Despite the differences, the Charlson index score derived from the administrative data had an identical ability of predicting in-hospital mortality to the score derived from chart data (5). Weiner and colleagues advocate a broad use of administrative data for monitoring quality and our uses fall within their recommendations (6). Romano and Mark assessed the sensitivity and reliability of coding for common diagnoses and procedures using California discharge abstracts and found in 7 of 8 comorbidity categories, sensitivity exceeded 85% (7). Weingart et. al. used administrative data, specifically a complications screening algorithm to identify inpatient complications using physician judgment as the gold standard and found flagged complications in 68.4% of surgical cases and 27.2% of medical cases (8). Yasmeen et. al. examined the sensitivity and positive predictive value to validate the coding of obstetric diagnoses and procedures in hospital-reported data using the medical record as the gold standard and found that surgical procedures and birth deliveries were accurately reported with sensitivities and PPVs exceeding 90% (9). Quam et.al. found that claims data that includes diagnostic and pharmacy data yields a high level of concordance with the medical record and survey data in the identification of a specific medical condition (10). Studies have shown high sensitivity of 72% and specificity of 95% for high risk conditions with overall accuracy of 90% obtained from administrative billing data among children with high-risk conditions including asthma which made up 87%

of the high risk conditions (11), and high predictive value among adolescents and adults with asthma (12). Twiggs et. al. found that the combined use of both medical and pharmaceutical claims was more effective in identifying asthmatics than either one by itself (13). HEDIS criteria using administrative data support peer reviewed research, for example in patients with persistent asthma based on HEDIS criteria in five Medicaid programs (Colorado, Georgia, Indiana, New Jersey, Washington) using ICD-9-CM code 493.x to measure filling prescriptions of asthma control medication and the ratio of controller medication to the total number of medication prescriptions filled within one year (14). Fowles and colleagues report sensitivity and specificity of claims compared with ambulatory medical records to identify asthma was 0.82 and 0.99, respectively. Sensitivity of .82 using claims was higher than sensitivity using self-report at .64 (15). Wilchesky compared chart abstraction to diagnoses obtained from administrative database: asthma claims were highly specific, Sp= 96.76 (95%CI 96.5, 97.0). Although sensitivity for most conditions was below 60%, sensitivity was enhanced when all claims for services were assessed, as we propose to do (16). Bronstein et al found that 88.3% of diagnoses asthma on claims agreed with medical record, with a negative predictive value of 0.85 and a positive predictive value of 0.88.They conclude that claims are generally an accurate indicator of the content of a patient encounter. (17) Steinwachs et al. compared billed claims to medical records based on date of visit and diagnosis, on average, 90% of billed visits were documented in the medical record, for asthma there was 90.9 percent of billed visits in record on same date and 82.8 percent of billed visits with same diagnosis in record on same date. (18) Quan et al documented the validity of ICD-9-CM and ICD-10 coding systems in coding clinical information and found that ICD-10 data was generally comparable with that of ICD-9-CM data in recording clinical information (19). Regarding our capacity to identify exclusions, Quan et al found that claims had a PPV of 91.9, and a negative predictive value of 92.6, with *k* of 0.65 (substantial agreement[1]) compared to chart review for chronic pulmonary disease . ICD 10 performed similarly in this study (19).

From a public health perspective, asthma surveillance systems in several states, including Maine, North Carolina, Connecticut and Michigan, have shown the feasibility of using administrative data to identify children having asthma, based on primary and secondary diagnosis codes reported on inpatient and outpatient claims. In addition to identifying asthma, important demographic data such as gender, race/ethnicity, program of enrollment and county of residence (urbanicity) can be used to assess associations between utilization services for asthma, including ED visits or hospitalizations, and demographic characteristics. Risk factor information from administrative data can be used to target educational programs, clinical assessments, and treatment programs (20-23).

Researchers also classified children with evidence of persistent asthma using HEDIS criteria, (24). Another study showed the usefulness of ICD9 493.x to identify asthma for a quality measure using Maryland Medicaid Claims data (25). Like our measure, those researchers excluded children with a diagnosis of cystic fibrosis (ICD9 277) (25). Schneeweiss commented that misclassification errors from claims data are asymmetric, with specificity typically exceeding 95% and sensitivity often less (26). Such a pattern makes it unlikely that an accountable entity would be held accountable for patients that do not actually have asthma.

As noted in 1.67 above, as part of an alpha test for our measure we used a contractor to survey more than a dozen hospitals across three CAPQuaM measure sets. Responses from 10 hospitals were specific to asthma. We found that variables including date of birth, race, ethnicity, county of residence, primary and secondary diagnosis of asthma in the ED, hospitalizations, payment source, and others were reported to be readily available and easy to access within the medical record.

In light of the literature review and our alpha test, we attest that the data elements for the measure match those assessed in the literature and our alpha test, with most being supported by both the literature review and the alpha test. We further note that our data element use is consistent with health care industry standards.

Validity of Measuring ED Visits and Exemplar Panel Findings

A national expert panel was convened and applied the RAND/UCLA appropriateness method to reviewing the constructs underlying this measure. The 9 member panel also supported the measure itself without objection.

[1] The *k* value indicates a near perfect agreement (*k*: 0.81-1.0 between coded data and chart review data), substantial agreement (*k*: 0.61-0.80), moderate agreement (*k*: 0.41-0.60), and fair agreement (*k:* 0.21-0.40).

As the constructs of this measure are defined via the expert panel process and this is an innovative approach to measuring undesirable asthma outcomes, there is no gold standard or statistical analysis. As an outcome measure, no association with process needs to be tested, although the NHLBI guideline discusses ED visits and hospitalizations as undesirable and potentially preventable outcomes. We used Median Scores from the panel ratings of at least 7 to identify desirable constructs for the measure.

Some interesting exemplar ratings are shown below, with some key findings bolded:

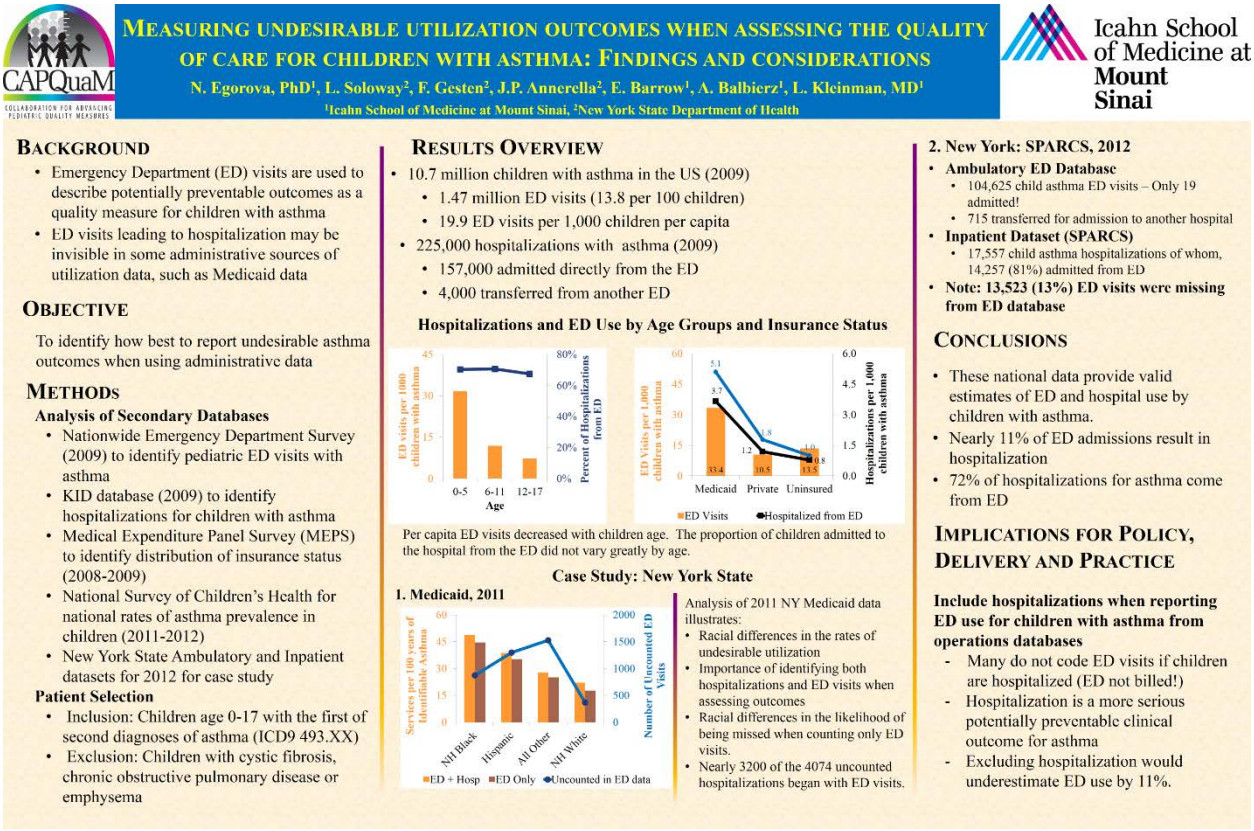
Scenario	MED
In general, this measure is intended to describe care for children who have asthma and identifiable since before the ED visit.	9
Asthma is established by a single prior hospital admission with asthma as the primary discharge diagnosis	9
A single admission is not sufficient to establish the presence of asthma.	1
In children after their 5th birthday, Asthma is established by a single prior ED visit with asthma as the primary discharge diagnosis	8
In children after their 5th birthday, Asthma is established by a single prior ED visit with asthma as the secondary discharge diagnosis	7
In children after their 5th birthday, Asthma is established by a single prior ED visit with asthma as any discharge diagnosis	6
In children after their 5th birthday, Asthma is established by 2 or more outpatient visits with asthma as a diagnosis.	9
In children after their 5th birthday, Asthma is not established until 4 or more outpatient visits with asthma as a diagnosis	2
Asthma related medication use helps to establish the presence of asthma.	**8**
Prescription for leukotriene inhibitors are typically asthma related..	**5**
Prescriptions for long acting beta 2 agonists are typically asthma related.	9
Prescriptions for inhaled steroids are typically asthma related.	8
Oral steroid bursts are typically asthma related.	5
In order to establish a diagnosis of asthma, a child should experience a total of at least 2 asthma related events such as outpatient visits for asthma and or asthma related prescriptions, one of which must be an outpatient visit	8
Filled prescriptions should not be considered when establishing the presence of asthma	1
Children with a diagnosis of COPD with chronic aspiration should be excluded from this measure.	**9**
Children with a diagnosis of COPD should be excluded from this measure.	**9**
Children with a diagnosis of cystic fibrosis should be excluded from this measure.	**9**
Children with a diagnosis of emphysema and chronic aspiration should be excluded.	**9**
Children with a diagnosis of emphysema should be excluded	**9**
The time frame for establishing a diagnosis of asthma extends before the reporting year.	**9**
This measure should include children over 2	8
The upper age limit for this measure should be children until their 20th birthday	4
The upper age limit for this measure should be children until their 21st birthday	**9**
For reporting purposes, adolescents 19-21 should be grouped with adolescents under 18.	5
For the purposes of this measure, only ED visits with asthma as the primary diagnosis are eligible for inclusion.	3
For the purposes of this measure, only ED visits with asthma as the primary or secondary diagnosis are eligible for inclusion.	**8**
For the purposes of this measure, all ED visits with asthma as a diagnosis are eligible for inclusion.	5
For the purposes of this measure, a treatment for asthma must be provided or prescribed in order for the ED visit to be eligible for inclusion.	3
In children prior to their 5th birthday, Asthma is established by 3 or more outpatient visits with asthma as a diagnosis	**9**
In children prior to their 5th birthday, Asthma is not established until 4 or more outpatient visits with asthma as a diagnosis	3

Our approach to identifiable asthma was validated by comparing the prevalence of identifiable asthma to the number of children with NY asthma claims, and to the prevalence estimate expected via analysis of the National Survey of Children's Health and to the prevalence of children with preventable asthma as defined by the NCQA's asthma measures. We sought to have a measure that would be much more inclusive than the persistent asthma criteria but still filtered with a threshold requirement. Indeed our findings supported this with more than 25% of all children with asthma claims eliminated by our definition, a denominator that was about 50% of the estimated survey-reported

lifetime incidence of asthma and 2.8 fold the number of children included than the NCQA criteria. We note that the NSCH survey prevalence exceeded the single claim approach.

Our own research looking at NY State Medicaid and national all payer data (see poster below, which was presented at peer-reviewed AcademyHealth national meeting) is consistent with expert and other recommendations that to identify all ED visits, one also needs to include hospitalizations for asthma as potential indicators of an otherwise unrecognized ED visit, which we have done and incorporated into the specifications.

In NY State using ED visits alone would miss about 13% of ED visits, nationally about 11%. The inclusion of hospitalizations will overestimate the number of ED visits by between 4 and 5 percent. As many of these hospitalizations are for acute exacerbations, the construct of undesirable utilization outcomes would include them, so that while the estimate is likely to be a bit high for ED visits, it is a fair estimate of asthma outcomes. Our approach to avoid de-duplication and double counting of an ED visit and its associated admission as two numerator events is specified (admission on same or next day in the same institution) in a manner that will slightly underestimate numerator events, thus compensating in part for the overestimation of ED events that may occur by including hospitalization.



REFERENCES

(1) Huzel, L. et al. Diagnosing Asthma: The fit between survey and administrative database. Canada Resp. Journal 2002.

(2) Jollis, J.G., M. Ancukiewicz, E.R. DeLong, D.B. Pryor, L.H. Muhlbaier, and D.B. Mark. 1993. ''Discordance of Databases Designed for Claims Payment versus Clinical Information Systems. Implications for Outcomes Research.'' Annals of Internal Medicine 119: 844–50.

(3) Lee, D.S., L. Donovan, P.C. Austin, Y. Gong, P.P. Liu, J.L. Rouleau, and J.V. Tu. 2005. ''Comparison of Coding of Heart Failure and Comorbidities in Administrative and Clinical Data for Use in Outcomes Research.'' Medical Care 43: 182–8.

(4) Muhajarine, N., C. Mustard, L.L. Roos, T.K. Young, and D.E. Gelskey. 1997. ''Comparison of Survey and Physician Claims Data for Detecting Hypertension.'' Journal of Clinical Epidemiology 50: 711–8.

(5) Quan, H., G. Parson, and W. Ghali. 2002. ''Validity of Information on Comorbidity Derived from ICD-9-CM Administrative Data.'' Medical Care 40: 675–85.

(6) Weiner, J.P., Powe,. R., Steinwachs, D.M., & Dent, G. (1990). Applying insurance claims data to assess quality of care: A compilation of potential indicators. Quality Review Bulletin, 16, 424-438.
(7) Romano, P.S., and D.H. Mark. 1994. ''Bias in the Coding of Hospital Discharge Data and Its Implications for Quality Assessment.'' Medical Care 32: 81–90.
(8) Weingart, S.N., L.I. Iezzoni, R.B. Davis, R.H. Palmer, M. Cahalane, M.B. Hamel, K. Mukamal, R.S. Phillips, D.T. Davies Jr., and N.J. Banks. 2000. ''Use of Administrative Data to Find Substandard Care: Validation of the Complications Screening Program.'' Medical Care 38: 796–806.
(9) Yasmeen, S., P.S. Romano, M.E. Schembri, J.M. Keyzer, and W.M. Gilbert. 2006. ''Accuracy of Obstetric Diagnoses and Procedures in Hospital Discharge Data.'' American Journal of Obstetrics and Gynecology 194: 992–1001.
(10) Quam, L., Ellis, L. B. M., Venus, P., Clouse, J., Taylor, C. G., & Leatherman, S. (1993). Using claims data for epidemiologic research: the concordance of claims-based criteria with the medical record and patient survey for identifying a hypertensive population. Medical Care, 31(6), 498-507.
(11) Daley MF, Barrow J, Pearson K, Crane LA, Gao D, Stevenson JM, et al. Identification and recall of children with chronic medical conditions for influenza vaccination. Pediatrics 2004;113(1 Pt 1):e26-33.
(12) Vollmer WM, O'Connor EA, Heumann M, Frazier EA, Breen V, Villnave J, Buist AS. Searching multiple clinical information systems for longer time periods found more prevalent cases of asthma. J Clin Epidemiol 2004;57:392-7.
(13) Twiggs, J. E., Fifield, J., Apter, A. J., Jackson, E. A., & Cushman, R. A. (2002). Stratifying medical and pharmaceutical administrative claims as a method to identify pediatric asthma patients in a Medicaid managed care organization. Journal of Clinical Epidemiology, 55(9), 938-944.
(14) Samnaliev, M., Baxter, J. D., & Clark, R. E. (2009). Comparative evaluation of two asthma care quality measures among Medicaid beneficiaries. Chest, 135(5), 1193-1196.
(15) Fowles, J. B., Fowler, E. J., & Craft, C. (1998). Validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. Journal of Ambulatory Care Management, 21(1), 24-34.
(16) Wilchesky, M., Tamblyna, R. M., & Huang, A. (2004). Validation of diagnostic codes within medical services claims. Journal of Clinical Epidemiology, 57, 131-141.
(17) Bronstein, J. M., Santer, L., & Johnson, V. (2000). The use of Medicaid claims as a supplementary source of information on quality of asthma care. Journal for Healthcare Quality, 22(6), 13-18.
(18) Steinwachs, D. M., Stuart, M. E., Scholle, S., Starfield, B., Fox, M. H., & Weiner, J. P. (1998). A comparison of ambulatory Medicaid claims to medical records: A reliability assessment. American Journal of Medical Quality, 13(2), 63-69.
(19) Quan, H., Li, B., Saunders, L. D., Parsons, G. A., Nilsson, C. I., Alibhai, A., et al. (2008). Assessing validity of icd-9-cm and icd-10 administrative data in recording clinical conditions in a unique duallycoded database. HSR: Health Services Research, 43(4), 1424.
(20) Deprez RD, Asdigian NL, Oliver LC, Anderson N, Caldwell E, Baggott LA. Development of a prototype system for statewide asthma surveillance. Am J Public Health 2002;92:1946-51.
(21) Buescher PA, Jones-Vessey K. Using Medicaid data to estimate state- and county-level prevalence of asthma among low-income children. Matern Child Health J 1999;3:211-6.
(22) Children's Health Council. Asthma affects nearly one in ten children enrolled in HUSKY Part A in 1999–2000. Hartford (CT): Children's Health Council; 2001.
(23) Dombkowski, K. J., Wasilevich, E. A., & Lyon-Callo, S. K. (2005). Pediatric asthma surveillance: Using Medicaid claims. Public Health Reports, 120, 515.
(24) National Committee for Quality Assurance. Use of appropriate medications for people with asthma. HEDIS 2003, volume 2: technical specifications. Washington: NCQA; 2003.
(25) Stuart, B., Singhal, P. K., Magder, L. S., & Zuckerman, I. H. (2003). How robust are health plan quality indicators to data loss? A monte carlo simulation study of pediatric asthma treatment. HSR: Health Services Research, 38(6), 1547.
(26) Schneeweiss, S., & Avorn, J. (2005). A review of uses of health care utilization databases for epidemiologic research on therapeutics. Journal of Clinical Epidemiology, 58(4), 323-337.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Please see the section above. The face validity of our expert panel, the test-test reliability, the critical importance of having a standard, reliable, and valid approach to measuring the rate of asthma ED visits all support this measure.

We interpret our measure to be a valid estimate of the rate of ED visits and an even better estimate of undesirable outcomes from asthma.

Our interpretation is that administrative data are reliable for identifying asthma, and that year to year test retest reliability seems to indicate similar patterns of performance when identifying ED visits for asthma, reinforcing the reliability of our operational definitions for identifying eligible children. Our specification provide a sensitive and face valid approach to identifying an unbiased sample of children with ED visits (ensuring we don't bias the results towards the inappropriate by missing those with hospitalization).

Most databases contain consistent elements, are available in a timely manner, provide information about large numbers of individuals, and are relatively inexpensive to obtain and use. Validity of many databases has been established, and their strengths and weaknesses relative to data abstracted from medical records and obtained via survey have been documented (30). Administrative data are supported, if not encouraged by federal agencies, such as NIH, AHRQ, HCFA, and the VA. The Centers for Medicare & Medicaid Services has made clear to the participating AHRQ-CMS CHIPRA Centers of Excellence funded to develop measures in the Pediatric Quality Measures Program that it places a premium on feasibility when assessing those measures that it will most highly recommend to states to complete. The sources of data for the existing measure and other similar measures are typically based upon administrative data as well, providing consensual validation for using administrative data as the primary data source.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b4*

Exclusions are clinical and specifically guided by the explicit criteria developed by the expert panel.

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Denominator Exclusions: Children with concurrent or pre-existing: Chronic Obstructive Pulmonary Disease (COPD) diagnosis (ICD-9 Code: 496), Cystic Fibrosis diagnosis (ICD-9 code 277.0, 277.01. 277.02, 277.03, 277.09), or Emphysema diagnosis (ICD-9 code 492xx). Children who have not been consecutively enrolled in the reporting plan for at least two months prior to the index reporting month, as well as the index reporting month itself.

There are no numerator exclusions.

Exclusions were only included if they were endorsed by the expert panel. In studying the denominator we found that a very few percent of potentially eligible children (<=2.5%) were excluded by clinical diagnoses. The use of three months of continuous enrollment was recommended by our multi-stakeholder consortium and avoids the exclusion of more than 20% of otherwise eligible children from the population with identifiable asthma compared to a 12 month requirement.

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

In order to develop a sample of approximately 125,000 children with asthma in our initial field test (that required a 12 month continuous enrollment criterion), we excluded 212 with COPD, 650 with cystic Fibrosis and 482 with emphysema

(those children were not mutually exclusive, in other words, children may have been excluded for more than one reason so the total number of exclusions was at least 212 and less than the sum of the three diagnoses (between 1.6% and 2.5% of otherwise eligible children).

Had we used a 12 month continuous enrollment criterion, we would have excluded more than 20% of otherwise eligible children.

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Exclusions are clinical and represent construct validity rather than statistical considerations.

The exclusions are purposeful and not statistical, and are based upon the findings of the expert panel. Noise is likely to be reduced by the exclusion of key diagnoses. Longer continuous enrollment requirements would harm validity since large number of children with real symptoms who are established and being managed for asthma would have been excluded. The 3 month continuous enrollment requirement is also conceptual, requiring the children to be under the management of the health plan in order to ask the plan to accept accountability.

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES
If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?
☐ **No risk adjustment or stratification**
☐ **Statistical risk model with** Click here to enter number of factors **risk factors**
☒ **Stratification by** 1 **risk categories**
☐ **Other,** Click here to enter description

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

Specifications for this measure requires stratification and reporting by age group only and also within age group by race/ethnicity. Several additional stratifications are optional but may be requested by the accountability entity or provided by the accountable entity. These variables include rurality/urbanicity and county level of poverty. Our findings suggest that risk adjustment is not critical for interpreting the results or for validity, but that stratification is informative to help to promote like to like comparisons and allow for plans to demonstrate how they do on specified subgroups. Such voluntary stratification specified in the measure helps to mitigate against the potential for misinterpretation and unintended consequences.

Within age group, we specify a number of stratifications as we have done for all of our CAPQuaM PQMP measure. Absent clear biological evidence that ED visits should be more likely in any of the sub categories we have chosen not to adjust but to report both topline and stratified results within age groups.

We used stratification to allow for a granular understanding of performance. Biological data and national guidelines agree and do not support risk adjustment to control for patient characteristics on the variables of interest. The Pediatric Quality Measures Program which funded development of this measure requests that measures be specified to be able to identify disparities and differences by a variety of characteristics and this measure does that.

The NIH NHLBI NAEPP guideline notes that goals of care and definition of successful management are the same regardless of baseline presentation. Hence clinical risk adjustment is not appropriate.

As indicated by the NHLBI guideline (http://www.nhlbi.nih.gov/files/docs/guidelines/asthgdln.pdf page 38)

> ***"An important point linking asthma severity, control, and responsiveness is that the goals are identical for all levels of baseline asthma severity.*** *A patient who has severe persistent asthma compared to a patient who has mild persistent asthma, or a patient who is less responsive to therapy may require more intensive intervention to achieve well-controlled asthma; however, the goals are the same: in well-controlled asthma, the manifestations of asthma are minimized by therapeutic intervention."*

For reasons other than controlling for case mix, we specify this measure to be stratified by age group and race/ethnicity as well as providing a top line analysis. Without such stratifications, racial and ethnic disparities (which have been found to be prevalent in children with asthma) might go unnoticed. The CHIPRA legislation that funded the development of these measures asks for the capacity to identify such disparities to be included in the measure specifications.

Specifications for further stratifications, such as by rurality/urbanicity and by county level of poverty are provided, in the event such stratification is requested by the accountability entity or desired by the reporting entity.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Consistent with the Disparities Working Group of the Pediatric Quality Measurement Program, CAPQuaM has chosen an approach to not risk adjust for outcomes as being most appropriate to measuring actual performance. Nonetheless, we honor the parameters in the legislation funding the PQMP and also recognize the interest of various stakeholders in comparing like-to-like: hence we have specified key stratifications for analysis and presentation. The accountability entity has the option to request the granularity of stratification that suits its needs beyond age strata and race/ethnicity.

The conceptual model is that of CAPQuaM that includes that in pediatrics age is a key predictor and stratification is valuable. We were asked by AHRQ and CMS to include other constructs and we have manifest them as specified, such as race/ethnicity, poverty level in the caregivers county of residence, rurality/urbanicity on the caregiver's county of residence, insurance type and plan type, when variable. We have not added a stratum for children with special health care needs since asthmatics going to the emergency room are highly likely to belong in this category.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

n/a

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b4.9. Results of Risk Stratification Analysis:

Data presented in chart and figure show asthma outcomes stratified by age and race/ethnicity. Additional analyses showed meaningful stratifications by time of year, county level of poverty, and rurality/urbanicity.

Rate by Age and Race/Ethnicity (per 100 child-years)

Age Group	Race/Ethnicity	Rate	N
2 to 6 years	Non-Hispanic Not-Black	19.7	28,559
	Non-Hispanic Black	47.6	11,305
	Hispanic	32.7	22,524
7 to 12 years	Non-Hispanic Not-Black	12.3	34,766
	Non-Hispanic Black	27.9	16,825
	Hispanic	20.5	30,391
13 to 18 years	Non-Hispanic Not-Black	11.7	23,587
	Non-Hispanic Black	22.0	11,240
	Hispanic	15.1	18,251
19 or 20 years	Non-Hispanic Not-Black	16.9	8,088
	Non-Hispanic Black	31.1	2,797
	Hispanic	20.3	4,096

Rate by Age Group (per 100 child-years)

Age Group	Rate	N
2 to 6 years	29.7	62,388
7 to 12 years	18.7	81,982
13 to 18 years	15.1	53,078
19 or 20 years	20.5	14,981

Rate by Race/Ethnicity

(per 100 child-years)		
Race/Ethnicity	**Rate**	**N**
Non-Hispanic Not-Black	14.4	95,002
Non-Hispanic Black	22.3	75,262
Hispanic	31.0	42,168

All of these are statistically significant by chi square analysis, p well below 0.05.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i*.e., what do the results mean and what are the norms for the test conducted*)

We acknowledge the association of the stratification variables with the rate of asthma ED visits but have not found evidence justifying such differences as either acceptable or un-modifiable by health care. Indeed there is evidence that primary care, adherence to guidelines, and other healthcare interventions can reduce or eliminated the impact of these factors. Federal guidelines quoted above support this perspective. We have expanded our requirement for stratification to create a "Both-And" presentation of the results and to enhance a more granular interpretation of the findings from this measure and to reduce the likelihood of misinterpretation or unintended consequences from measurement.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Contingency table analysis with chi-square and using t-statistics were coherent and each illustrated the presence of statistical differences. Additionally, Poisson Regression analyses indicate significant differences by health plans and by counties, whether or not controlling for age group and race/ethnicity. As is desirable, the number of significant differences demonstrated depend upon which of the entities is used as a comparator in the analysis. Using a typical plan/county results in a smaller proportion of significant differences than selecting a more extreme plan/county.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

We have described ZIP models and their results in the sections of 2a2. Here we describe a complementary analysis using Poisson.

Differences between major subgroups were statistically significant, including race/ethnicity, age group, level of poverty in the county, and level of urbanicity (urban, suburban, rural). All one way and two way (within age stratum) chi square analyses and t test analyses were $p<0.05$.

We further note that the measure was sensitive enough to demonstrate face validity with statistically significant differences from month to month and season to season as expected for this outcome. Within the NY State Medicaid

data, differences were also found by eligibility category, which in this case can serve as proxy for health plan. Chi Square of the rate difference (7.7 ED visits per 100 children) between those qualifying for cash assistance versus those qualifying because of SSI was 32.07 with one degree of freedom meaningfully exceeds the critical value of 10.828 for p<.001. This demonstrates excellent capacity to distinguish between health plans.

Specifically, within the NY State Medicaid data, there are 22 plans identified. Using a poisson regression analyses, we found that 18 Managed Care Organization plans that included at least 900 children contributing time to the denominator yielded statistically significant differences among plans and among counties, whether or not we controlled for age group and/or race and ethnicity and/or urbanicity. This is true also when we analyzed stratified by age group.

8 of these plans were fully managed and 10 were partially managed plans. Partially managed plans had a statistically higher rate than fully managed HMO plans. .Among the 18 plans, the mean rate is 15.7, with a standard deviation of 6.0. A SAS summary of the rate distribution is shown below.

Quantiles (Definition 5)	
Level	Quantile
100% Max	30.03732
99%	30.03732
95%	30.03732
90%	28.70196
75% Q3	17.62075
50% Median	13.65690
25% Q1	11.67487
10%	9.19899
5%	9.07094
1%	9.07094
0% Min	9.07094

The exemplar table below shows when a randomly selected plan is chosen as the comparator in a model that controls for Black race, Hispanic ethnicity, and age group. Highlighted rows are statistically different from the index plan, demonstrating the power to for signal to be interpreted over noise, even when controlling for the above factors. Uncontrolled analyses produced analogous results.

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > ChiSq
Intercept		1	-4.4112	0.0303	-4.4704 -4.3519	21264.0	<.0001
plan	1	1	-0.1376	0.0294	-0.1952 -0.0800	21.94	<.0001
plan	2	1	-0.1701	0.0386	-0.2458 -0.0944	19.39	<.0001
plan	3	1	0.1310	0.0274	0.0772 0.1848	22.76	<.0001
plan	4	1	-0.4571	0.0730	-0.6001 -0.3140	39.23	<.0001
plan	5	1	-0.4727	0.0479	-0.5665 -0.3789	97.56	<.0001
plan	6	1	0.0402	0.0265	-0.0117 0.0921	2.31	0.1287
plan	7	1	0.4882	0.0218	0.4455 0.5310	501.20	<.0001
plan	8	1	-0.0725	0.0733	-0.2162 0.0713	0.98	0.3232
plan	10	1	-0.2006	0.0551	-0.3086 -0.0925	13.22	0.0003
plan	11	1	-0.3049	0.0508	-0.4043 -0.2054	36.07	<.0001
plan	12	1	-0.3607	0.1098	-0.5760 -0.1455	10.79	0.0010
plan	13	1	-0.2146	0.0629	-0.3380 -0.0913	11.63	0.0006
plan	14	1	-0.3115	0.0722	-0.4529 -0.1700	18.62	<.0001
plan	15	1	-0.6056	0.0375	-0.6791 -0.5320	260.63	<.0001
plan	16	1	-0.5285	0.0587	-0.6435 -0.4136	81.15	<.0001
plan	17	1	0.4491	0.0213	0.4074 0.4908	445.49	<.0001
plan	21	1	-0.4093	0.0807	-0.5675 -0.2511	25.72	<.0001
plan	22	0	0.0000	0.0000	0.0000 0.0000	.	.
black		1	0.6985	0.0160	0.6671 0.7299	1902.69	<.0001
Hisp		1	0.2927	0.0154	0.2625 0.3229	360.50	<.0001
agegrp	1	1	0.3165	0.0268	0.2639 0.3691	139.06	<.0001
agegrp	2	1	-0.2037	0.0268	-0.2563 -0.1511	57.64	<.0001
agegrp	3	1	-0.4135	0.0285	-0.4694 -0.3576	210.25	<.0001
agegrp	4	0	0.0000	0.0000	0.0000 0.0000	.	.
Scale		0	1.0000	0.0000	1.0000 1.0000		

Note: The scale parameter was held fixed.

County by county results also varied significantly and were aggregated by urban influence code and level of poverty and continued to show capacity to show meaningful differences. The distribution of findings across 45 counties is shown below (unadjusted) (45 counties with at least 1000 months contributed to the denominator) the mean is 10.9 (std. 7.9).

Quantiles (Definition 5)	
Level	Quantile
100% Max	53.33333
99%	53.33333
95%	19.13721
90%	16.81866
75% Q3	12.68946
50% Median	9.02256
25% Q1	7.55814
10%	4.56100
5%	2.60530
1%	1.64249
0% Min	1.64249

Using a randomly chosen rural county, there were a number of significant differences county to county. Selecting New York City as the comparator, all other counties differed significantly. This nuance in findings confirms the validity and specificity of differences as likely to be meaningful. We note that NYS analyses frequently contrast findings with those in NYC and meaningful differences are the rule and not the exception.

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

The measures are sensitive enough to detect meaningful differences as observed within a population (as with the seasonal and month to month variations described above) and across populations such as counties or plans, as we have demonstrated. They are sufficiently robust and precise to measure real differences and not to create artificial ones. We interpret our findings in aggregate to indicate that the signal to noise ratio is very strong for this measure. There is high certainty and confidence that the performance measure scores are both reliable and valid.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was*

used)

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

We use administrative claims to establish eligibility. No missing data analysis performed as we were using standard data sources that are contractually obligated to be provided to NYS Medicaid. We had a total of three children in our analysis of children with identifiable asthma who dropped out of the analyses because of any missing data element.

Our analyses found that the absence of pharmacy data would reduce only slightly (as we recall, less than 1%) the number of children identified as having identifiable asthma. This finding became apparent during alpha testing of our specifications and was incorporated into our specifications as a permissive allowance when pharmacy data were not available. We have not located the original analysis and hope for the NY State team to replicate the analysis by the time of the Committee meeting.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Addressed elsewhere regarding data sources and definition of identifiable asthma, requirements for 3 months of continuous enrollment. The use of a composite requirement to establish eligibility reduces that likelihood of systematic error or dependence upon any specific data field. The use of complementary sources of identifying visits (CPT codes and revenue codes) accomplishes a similar goal.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Generally N/A. Systems unable to integrate pharmacy data into the eligibility analysis would have a minimally higher risk population than those with pharmacy claims. The specifics of the definitions and the limited impact of pharmacy claims on eligibility combine to make the expected impact of this on the rate of ED visits to almost zero. They are included in the identification of denominator because our expert panel directed us to do so. More importantly, as cited above, the NHLBI guideline tells us that outcomes should not be adjusted for baseline risk, so this does not truly disadvantage a reporting entity according to the guideline. Further, unlike many asthma measures where the absence

of pharmacy data would systematically disadvantage identification of satisfactory performance, for this measure, pharmacy data is used only to complement other utilization data when determining eligibility. In theory there is likely to be a short period of time when a child would be identified using the pharmacy data and not by the other utilization criteria prior to meeting other 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.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Other
If other: In rare instances, race/ethnicity and zip code may need to be abstracted from the chart.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.
ALL data elements are in defined fields in a combination of electronic sources

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.
Medicaid systems typically include race and ethnicity and zip code as defined electronic fields.

Emergency department visits that result in hospitalization are often not coded as ED visits in administrative data. The most valid estimate of ED visit rate requires use of both ED and hospitalizations as numerator events.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None at present

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Not in use	

4a.1. For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

4a.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

We are awaiting NQF endorsement for use. There are no policies or actions of the developer/steward or accountable entities that would restrict access to performance results or impeded implementation.

The measure is not currently in use but its application is currently being explored by the UCSF Center of Excellence in the Pediatric Quality Measurement Program (Round 2).

The topic of ED asthma use was assigned to our measure development project in the Pediatric Quality Measures Program by CMS, by far the largest single third party payer for medical care for children in the US, and by AHRQ. This measure has received the imprimatur of the American Academy of Pediatrics as one of its high priority measures that emerged from their joint (with the American Board of Pediatrics) Measurement Alignment and Strategic Selection Work Group.

4a.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

The measure is a straightforward approach to estimating the rate of Emergency Department visit Use for children managed for identifiable asthma. Our analyses in NY State Medicaid data confirmed feasibility, usability, and responsiveness of the measure to substantive constructs including race/ethnicity, age group and by health plans. We find these data and their consistency with expected findings to be persuasive that the measure is both valid and sensitive to real differences. Therefore, when this measure is endorsed by NQF, it will be applicable to a variety of settings and organizations.

As a part of our work with PQMP, we are working on specific plans for dissemination and use. Our plan for implementation includes submitting our application for measurement endorsement from the National Quality Forum. We are having conversations with NY State Medicaid (who was one of our partners in development) regarding the application and use of this measure. No time frames have been established.

Meeting the expected timeframes of NQF, CAPQuaM intends for the measure to be used for an accountability application within 3 years of initial endorsement and public reporting within six years of initial endorsement.

At this point in time, the submitted measure has received the imprimatur of the American Academy of Pediatrics as one of its high priority measures that emerged from their joint (with the American Board of Pediatrics) Measurement Alignment and Strategic Selection Work Group.

Improvement
Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
Rates of emergency department visit use for children managed for identifiable asthma is an important outcome measure with intrinsic value that helps ensure high-quality, efficient healthcare for individuals and populations. Not all ED visits for asthma are necessary, some cases require a different level of care for the clinical circumstance. Also, a significant proportion of visits potentially could have been prevented with better prior management. A variety of stakeholders benefit from this measure:

- Plans could provide clinicians this data to use to accurately identify patients who benefit from enhanced asthma care.

- Health systems can use this data to distinguish patients who have identifiable asthma, their demographics, and the care they receive and the associated costs. This information allows practices, groups and facilities to evaluate and compare treatment plans between practice sites, medical and other professional groups and between integrated or other delivery networks. This evidence-based evaluation promotes the adoption of more effective and efficient health systems.

- States and healthcare agencies can also use these measures to compare larger systems to test and evaluate treatment options, payment models (e.g. managed care, primary care case management), quality of health plans, costs and health outcomes. Findings can be stratified by state or regionally (e.g. urban, rural, health shortage regions) to understand policy, demographic and culture effects.

- The data also allows clinical, public health and epidemiology researchers to understand the type, level and cost of care patients are receiving related to their health outcomes, giving opportunity to compare between health plans, payment models and treatment options. It also gives a deeper understanding of how individual (e.g. sex, age, gender, race, social economic status, poverty) and community determinants (e.g. work environment, community benefits) affect the rate of ED visits over time. Furthermore, these measures gives researchers the tools to identify and reach out to patients and their families to understand their health culture and practices to ensure that health services offered will most likely be utilized.

4c. Unintended Consequences
The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**
There has not been any evidence of unintended negative consequences to individual or populations. There are no anticipated unintended consequences if measuring at the level of comparing states, geographic regions, payment models, or health plans.

Comparing individual health care professionals is not recommended as care is provided across practices may be necessary. Also, it is not appropriate for a single hospital comparison because it is measuring the system performance not the hospital performance. Lastly, although the measure can be used to compare practice sites, medical or other professional groups or integrated or other delivery networks, the measures are only recommended for large practices or integrated delivery systems that own their own risk and manage inpatient and outpatient care or that have access to all payer data sources

4c.2. **Please explain any unexpected benefits from implementation of this measure.**

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.

4d2.2. **Summarize the feedback obtained from those being measured.**

4d2.3. **Summarize the feedback obtained from other users**

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization of Related Measures
The measure specifications are harmonized with related measures;
OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications harmonized to the extent possible?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
Our definition of identifiable asthma is more inclusive than, for example, NCQA's persistent asthma construct. We use similar medication definitions as NCQA, except we exclude leukotriene inhibitors from asthma-related medications because our expert panel felt that these medications were used frequently for allergy patients and judged that the small gain in sensitivity of identifying children (considering all criteria) would be less than the loss in sensitivity and likelihood to include non-asthmatic

children with allergies. Our specifications have been validated by an expert panel in the context of a peer reviewed process commissioned by AHRQ and CMS to advance the field and science of pediatric quality measurement beyond the state represented in pre-existing measures. The specification of a person-time denominator allows for the measure to have a shorter requirement for continuous enrollment than other measures with less risk of bias than previous measures.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Appendix_Asthma_rate_12_12_16.docx

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** University Hospitals Cleveland Medical Center
Co.2 **Point of Contact:** Lawrence, Kleinman, drlarrykleinman@gmail.com, 216-286-6969-
Co.3 **Measure Developer if different from Measure Steward:** University Hospitals Cleveland Medical Center
Co.4 **Point of Contact:** Lawrence, Kleinman, drlarrykleinman@gmail.com, 216-286-6969-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
Role: Expert Panelists
Elizabeth Allen, M.D Ohio State University
Chitra Dinakar, M.D. Department of Pediatrics University of Missouri-Kansas City
Stephen Teach, M.D., M.P.H. Children's National Medical Center
Charles Macias M.D., M.P.H. Pediatrics, Baylor Texas Children's Hospital
Michael Cabana M.D., M.P.H. University of California, San Francisco

Barbara Yawn M.D., M.S. Olmstead Medical Center, University of Minnesota
Joan Connell M.D. University of North Dakota
Delaney Gracy M.D., M.P.H. The Children's Health Fund and Montefiore Children's Hospital
Sharlene Miner M.D. Emergency Medicine Inter Mountain Medical Center

ROLE: Steering Committee and Investigator
Wilson Pace, MD American Academy of Family Physicians – DARTNET Institute - University of Colorado
Lynn Olson, PhD American Academy of Pediatrics
Christina Bethell, PhD, MBA, MPH Child and Adolescent Health Measurement Initiative, Johns Hopkins University (Previous OHSU)
Elizabeth Howell, MD Icahn School of Medicine at Mount Sinai
Harold Kaplan, MD Icahn School of Medicine at Mount Sinai
Lawrence Kleinman, MD, MPH Icahn School of Medicine at Mount Sinai

Rebecca Anderson Mount Sinai Medical Center
Ira Nash, MD Northshore - Long Island Jewish Medical Center (previous Mount Sinai School of Medicine)
Eyal Shemesh, MD Icahn School of Medicine at Mount Sinai
Mary Barton, MD National Committee on Quality Assurance
Charles Homer, MD, MPH US Department of HHS (previous National Institute for Child Health Quality)
Marla Clayman, PhD American Institutes for Research (previous Northwestern University)
Foster Gesten, MD New York State Dept. of Health, Office of Health Insurance Programs
Jerod M. Loeb, PhD The Joint Commission
Robert Rehm National Committee for Quality Assurance
Steve Kairys, MD American Academy of Pediatrics/QuIIN
Erin DuPree, MD The Joint Commission (previous: Icahn School of Medicine at Mount Sinai)
David Baker The Joint Commission (Previous Northwestern)
Beverley Johnson, BSN* Institute for Patient- and Family-Centered Care
Doris Peter Consumers Union

ROLE: Senior Advisory Board Member and Investigator
Shoshanna Sofaer, DrPH American Institutes for Research (previous CUNY Baruch)
Lynne Richardson, MD Icahn School of Medicine at Mount Sinai
Ian Holzman, MD Icahn School of Medicine at Mount Sinai
Marilyn Kacica, MD New York State Dept. of Health, Division of Family Health

ROLE: Senior Advisory Board Member
Marc Lashley, MD Allied Pediatrics
Gary Mirkin, MD Allied Pediatrics
Barbara Kupferman, RN, CMCN, LNC AmeriChoice by United Healthcare
L. Gregory Pawlson, MD, MPH SE Healthcare QUALITY Consultants (previous Blue Cross Blue Shield)
John Santa, MD* (previous Consumers Union)
John Clarke, MD ECRI, PA Patient Safety Authority
Scott Breidbart, MD Empire Blue Cross Blue Shield/ Anthem
Jeff Terry, MBA GE Healthcare
Robert St. Peter, MD, MPH Kansas Health Institute
Ruth Stein, MD Montefiore Children's Hospital
Arthur Aufses, MD Icahn School of Medicine at Mount Sinai
Eric Rose, MD Icahn School of Medicine at Mount Sinai
Wendy Brennan, MS National Alliance on Mental Illness of New York City (NAMI - NYC Metro)
Laurel Pickering, MPH Northeast Business Group on Health
Tony Hope, BM BCh, PhD Oxford University, The Ethox Center
Martin Hatlie, JD Partnership 4 Pt. Safety, Consumers Advancing Patient Safety
Paul Wise, MD Stanford University
Mark Chassin, MD, MPH, MPP The Joint Commission
Ira Nash, MD Icahn School of Medicine at Mount Sinai

ROLE: Investigators/Key Staff
Rusty McLouth, MS AAFP
Kasey McCracken, MPH CAHMI
Elise Barrow, MPH Icahn School of Medicine at Mount Sinai
Kasey Coyne, BA Icahn School of Medicine at Mount Sinai
Natalia Egorova, PhD Icahn School of Medicine at Mount Sinai
Elizabeth Howell, MD Icahn School of Medicine at Mount Sinai
Harold Kaplan, MD Icahn School of Medicine at Mount Sinai
Melissa Manice, MPH Icahn School of Medicine at Mount Sinai
Joshua Oppenheimer, BA Icahn School of Medicine at Mount Sinai
Eve Oster, MD Pediatric Partners of Gwinnett (previous Icahn School of Medicine at Mount Sinai)
Daniel O'Sullivan, BS University of Virginia (previous Icahn School of Medicine at Mount Sinai)
Michael Parides, PhD Icahn School of Medicine at Mount Sinai

Barbara Rabin, MHA	Icahn School of Medicine at Mount Sinai
Carolyn Rosen, MD	Icahn School of Medicine at Mount Sinai
Melissa Saperstein, MSW	Icahn School of Medicine at Mount Sinai
Eyal Shemesh, MD	Icahn School of Medicine at Mount Sinai
Andrew Ting, MD	Icahn School of Medicine at Mount Sinai
Adam Vella, MD	Icahn School of Medicine at Mount Sinai
Virginia Walther, MSW	Icahn School of Medicine at Mount Sinai
Jennifer Zeitlin, SD	Icahn School of Medicine at Mount Sinai, French Institute of Health and Medical Research (INSERM)
Marianne McPherson, PhD, MS	NICHQ
Joseph Anarella, MPH	NYS DOH, OHIP
Lee Sanders, MD, MPH	Stanford University
Kasey Coyne	Northwestern
Victoria Wagner	New York State Health Department Office Quality Patient Safety
Laura Soloway	New York State Health Department Office Quality Patient Safety
Kevin Malloy	New York State Health Department Office Quality Patient Safety
Judy Stribling	Icahn School of Medicine at Mount Sinai
Linda Radecki	AAP
Keri Thiessen	AAP – QuIIN
Laura Morris	New York State Department of Health
Olivia Lindly	OHSU
Kasey McCracken	OHSU
Sandeep Sharma	Icahn School of Medicine at Mount Sinai
Ann Nevar	UH Rainbow Babies and Children's Hospital/ CAPQuaM
Suzanne Lo	UH Rainbow Babies and Children's Hospital/ CAPQuaM
Amy Balbierz	Icahn School of Medicine at Mount Sinai
Samantha Raymond	Icahn School of Medicine at Mount Sinai
Allisyn Vachon	Icahn School of Medicine at Mount Sinai
Shannon Weber	Icahn School of Medicine at Mount Sinai
ROLE: Contractors	
Mary Fialkowski, MBA	The Joint Commission
Linda Hanold, MHSA	The Joint Commission
Elvira Ryan, MBA, BSN, RN	The Joint Commission
Sharon Sprenger, RHIA, CPHQ, MPA	The Joint Commission
Ann Watt, MBA, RHIA	The Joint Commission
Amy Slott	The Joint Commission
Denise Krusenoski	The Joint Commission
Elvira Ryan	The Joint Commission
Tasha Mearday	The Joint Commission

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:**

Ad.3 **Month and Year of most recent revision:**

Ad.4 **What is your frequency for review/update of this measure?**

Ad.5 **When is the next scheduled review/update for this measure?**

Ad.6 **Copyright statement:**

Ad.7 **Disclaimers:**

Ad.8 **Additional Information/Comments:**

Appendix 1: Tables and Figures

Table of Contents

Figure 1. Eligibility Algorithm Diagram

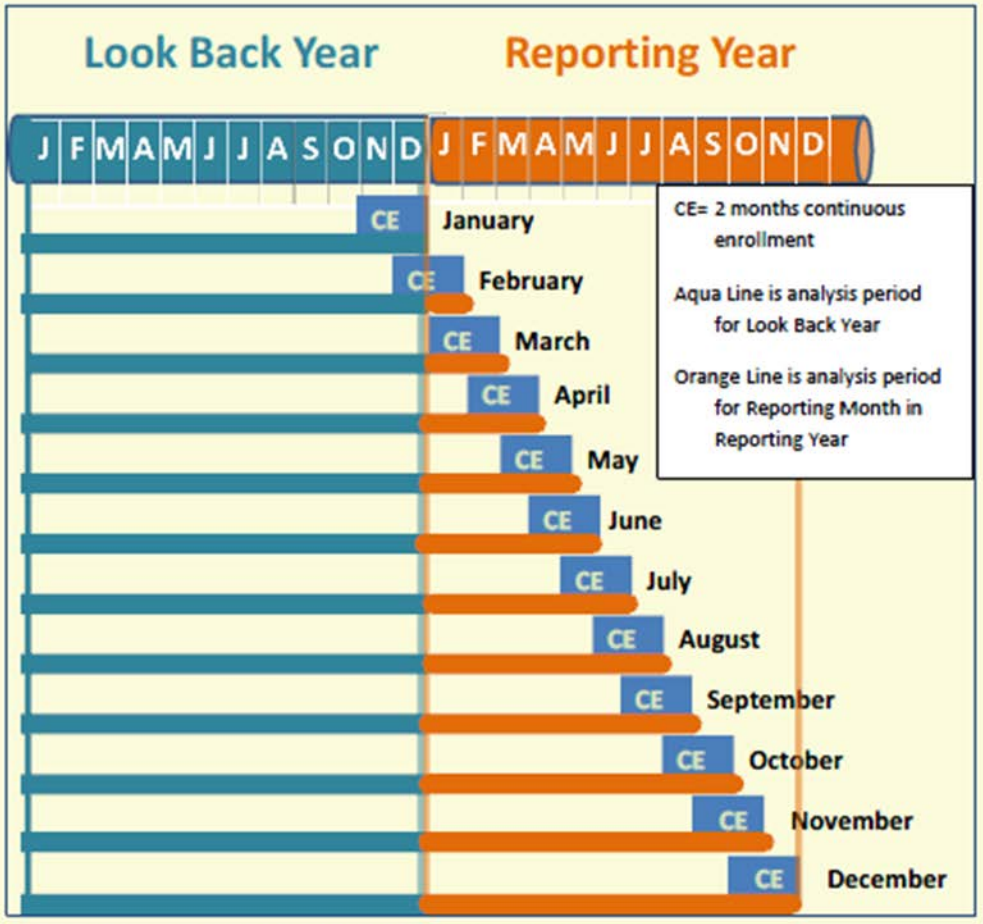


Diagram of Eligibility Requirements for Asthma Measure

Description of Diagram Components
Aqua: Look Back Year
Orange: Reporting Year
Blue: Continuous Enrollment (CE) period must be met to be eligible in a given reporting month
Black: Label of Reporting Month

Description of Eligibility Criteria Assessment
Eligibility Criteria are assessed month by month, using both the entire Look Back Year and the Reporting Year.

For any given reporting month, assess eligibility on 2 criteria. Eligible children are those that meet both of the following:

1) Has the child been continuously enrolled (CE) for the two months prior to the Reporting Month (3 months of continuous enrollment including the current Reporting Month)
2) Does the child meet the criteria for having been managed for persistent asthma in the assessment period (Aqua + Orange Lines)

Table 1. Month by Month Data, Stratified. New York State Medicaid Managed Care, 2012 In ED visits per 100 child-years.

Figure 2. Asthma ED Visits By Age and Month.

	2 - 5 y.o.	6-11 y.o.	12-18 y.o.	19-21 y.o.		Urban 1	Urban 2	Rural 1	Rural 2		Black	White	Hispanic
Jan	55								13.4	Jan	44.9	15.4	38.3
Feb	56								18.0	Feb	49.2	16.7	36.7
Mar	58								18.9	Mar	52.1	17.8	41.0
Apr	51								21.2	Apr	41.0	14.1	33.0
May	50								12.2	May	59.7	16.6	40.2
Jun	32.8	17.0	15.6	27.8	Jun	35.6	21.9	27.4	11.2	Jun	29.1	9.5	21.1
Jul	27.1	12.5	14.5	28.1	Jul	30.4	14.1	7.2	7.1	Jul	23.4	8.5	16.6
Aug	25.2	12.1	15.2	27.8	Aug	28.0	14.0	20.9	6.0	Aug	23.9	8.6	15.9
Sep	45.8	22.6	22.2	30.5	Sep	49.3	35.4	33.5	16.7	Sep	39.2	14.8	28.5
Oct	49.8	29.0	26.1	35.3	Oct	56.4	25.9	9.5	8.7	Oct	49.3	14.9	34.1
Nov	60.9	33.2	23.8	35.7	Nov	69.0	27.0	17.6	19.3	Nov	50.1	17.1	40.0
Dec	58.7	30.5	22.6	31.0	Dec	66.1	30.3	16.5	17.3	Dec	45.3	14.5	39.6

Asthma ED Visits By Age and Month

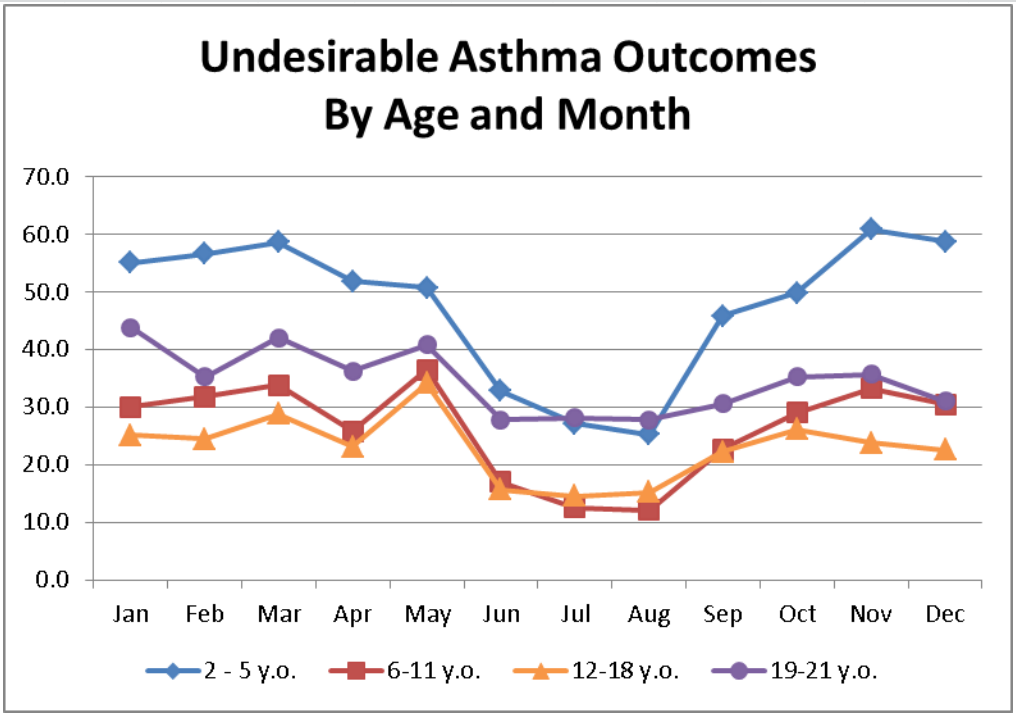


Figure 3. ED Visits per 100 Child-years by Age and Urbanicity

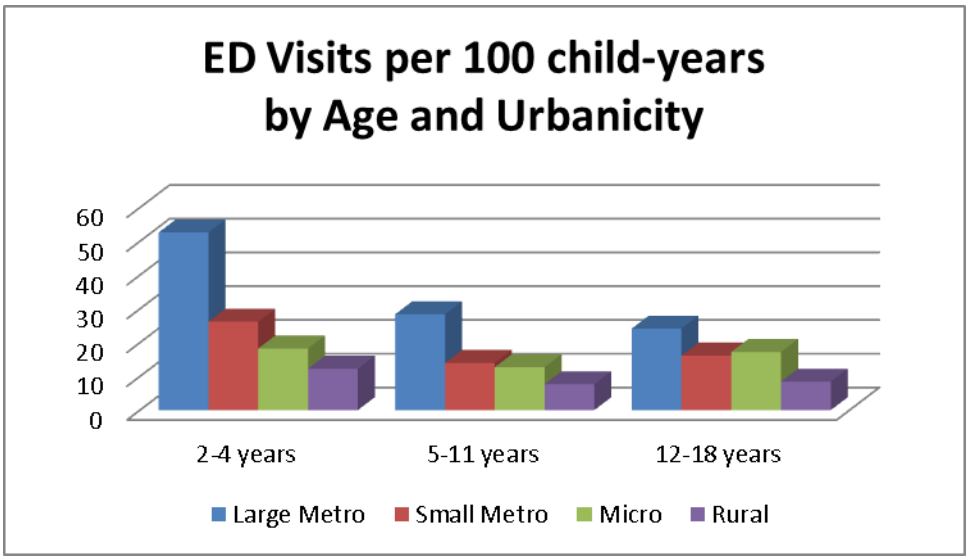


Figure 4. ED Visits per 100 child-years by Age and County Poverty quartile

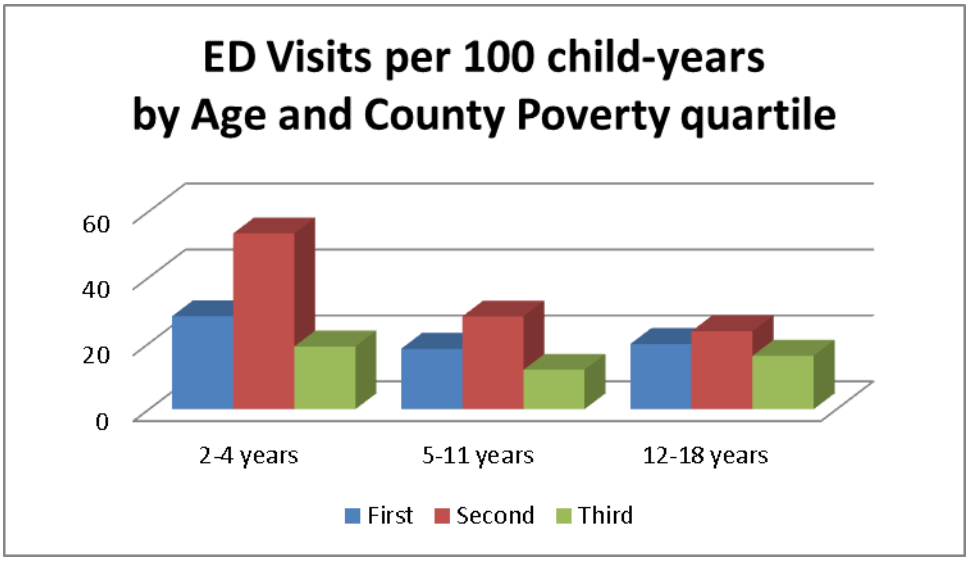


Figure 5. ED Visits per 100 Child Years by Age and Race/Ethnicity

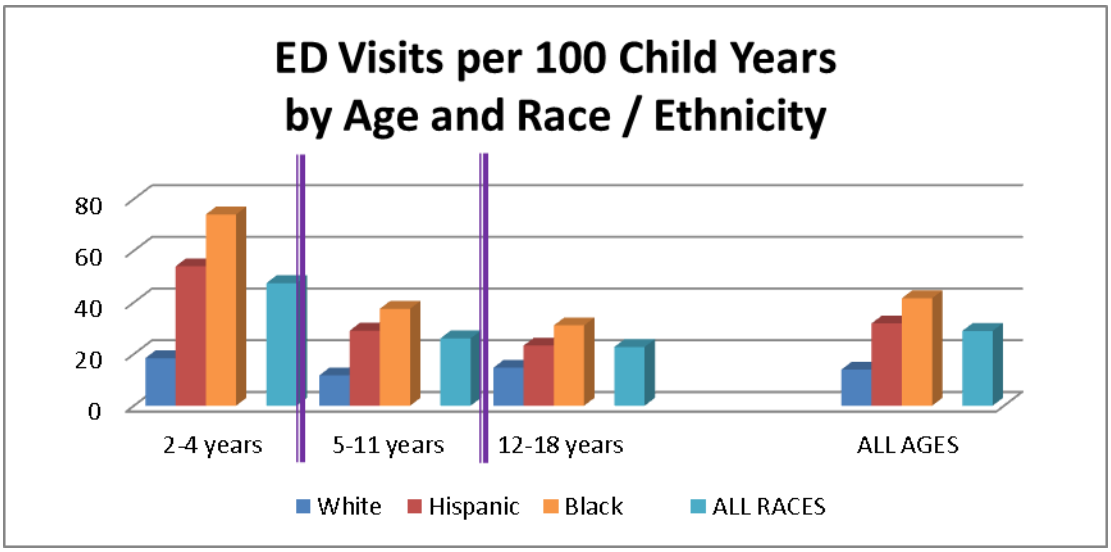


Table 3. ED Visits per 100 Child-years by Age and Quartile of Poverty

	White	Hispanic	Black		ALL RACES
2-4 years	18	54	74		47.44
5-11 years	12	29	38		26.03
12-18 years	15	23	31		22.74
ALL AGES	13.94	31.87	41.60		28.95

In visits per 100 child years.

Table 4. ED Visits per 100 Child-years by Age and Urbanicity

	Large Metro	Small Metro	Micro	Rural
2-4 years	52.6	26.2	18.3	12.3
5-11 years	28.4	14.0	12.7	7.6
12-18 years	24.1	16.2	17.2	8.4

Table 5. ED Visits per 100 Child-years by Age and Quartile of Poverty (First is lowest level of poverty)

Quartile of Poverty

	First	Second	Third
2-4 years	28	53	19
5-11 years	18	28	12
12-18 years	20	24	16

Table 6

Source of Systematic Review: • **Title**	

• **Author** • **Date** • **Citation, including page number** • **URL**	• National Heart, Lung, and Blook Institute, National Institutes of Health (NHLBI/NIH) Asthma Guideline 2007 www.nhlbi.nih.gov/guidelines/asthma (NAEPP Guideline) • http://www.nhlbi.nih.gov/health-pro/resources/lung/naci/asthma-info/asthma-guidelines.htm
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Quick Reference Guide: Asthma control focuses on two domains: 1)reducing impairment --- the frequency and intensity of symptoms… and 2) reducing risk – the likelihood of future asthma attacks… [later described as "prevent exacerbations] At the population level ED visits and hospitalizations represent failures of asthma control. Asthma Guidelines: - **Following science-based guidelines *works*** - Not only do they have the potential to improve a patient's *quality* of life; they can potentially *save a life*. **National asthma guidelines have been updated** In 2007, the National Asthma Education and Prevention Program (NAEPP), coordinated by the National Heart, Lung, and Blood Institute (NHLBI), released its third set of clinical practice guidelines for asthma. The Expert Panel Report 3—Guidelines for the Diagnosis and Management of Asthma (EPR-3) reflects the latest scientific advances in asthma drawn from a systematic review of the published medical literature by an NAEPP-convened expert panel. It describes a range of generally accepted best-practice approaches for making clinical decisions about asthma care. The EPR-3 emphasizes the importance of asthma control and focuses on two domains—current impairment and future risk—by which to assess asthma severity (for initiating therapy) and asthma control (for ongoing monitoring). EPR-3 also includes an expanded section on childhood asthma (with an additional age group), new guidance on medications, new recommendations on patient education in settings beyond the physician's office, and new advice for controlling environmental exposures that can cause asthma symptoms. **Asthma can be controlled** Scientific evidence clearly shows that most people could control their asthma by following current asthma clinical practice guidelines. With proper care, people who have asthma can stay active, sleep through the night, and avoid having their lives disrupted by asthma attacks. As a general rule, patients with well-controlled asthma should have: • Few, if any, asthma symptoms. • Few, if any, awakenings during the night caused by asthma symptoms. • No need to take time off from school or work due to asthma. • Few or no limits on full participation in physical activities. • No emergency department visits. • No hospital stays. • Few or no side effects from asthma medicines.

KEYPOINTS: OVERVIEW OF MEASURES OF ASTHMA ASSESSMENT AND MONITORING (pg. 36)

- The functions of assessment and monitoring are closely linked to the concepts of severity, control, and responsiveness to treatment:
    - Severity: the intrinsic intensity of the disease process. Severity is measured most easily and directly in a patient not receiving long-term-control therapy.
    - Control: the degree to which the manifestations of asthma (symptoms, functional impairments, and risks of untoward events) are minimized and the goals of therapy are met.
    - Responsiveness: the ease with which asthma control is achieved by therapy.
- Both severity and control include the domains of current impairment and future risk:
    - Impairment: frequency and intensity of symptoms and functional limitations the patient is experiencing or has recently experienced
    - Risk: the likelihood of either asthma exacerbations, progressive decline in lung function (or, for children, reduced lung growth), or risk of adverse effects from medication

KEY DIFFERENCES FROM 1997 AND 2002 EXPERT PANEL (pg. 37)
REPORTS

- The key elements of assessment and monitoring are refined to include the separate, but related, concepts of severity, control, and responsiveness to treatment. Classifying severity is emphasized for initiating therapy; assessing control is emphasized for monitoring and adjusting therapy. Asthma severity and control are defined in terms of two domains: impairment and risk.
- The distinction between the domains of impairment and risk for assessing asthma severity and control emphasizes the need to consider separately asthma's effects on quality of life and functional capacity on an ongoing basis (i.e., in the present) and the risks it presents for adverse events in the future, such as exacerbations and progressive loss of pulmonary function. These domains of asthma may respond differentially to treatment.

… p.38

An important point linking asthma severity, control, and responsiveness is that the goals are identical for all levels of baseline asthma severity. A patient who has severe persistent asthma compared to a patient who has mild persistent asthma, or a patient who is less responsive to therapy may require more intensive intervention to achieve well-controlled asthma; however, the goals are the same: in well-controlled asthma, the manifestations of asthma are minimized by therapeutic intervention.

	… page 41 regarding identification asthma, one key factor is: **The Expert Panel recommends that the clinician trying to establish a diagnosis of asthma should determine that (EPR–2 1997):** • **Episodic symptoms of airflow obstruction are present.** This is consistent with how we defined identifiable asthma… Page 63 It is important to evaluate the frequency, rate of onset, severity, and causes of exacerbations…severe exacerbations leading to ED visits and hospitalizations (Adams et al. 2000; Eisner et al. 2001; Ford et al. 2001; Lieu et al. 1998).
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	The National Asthma Education and Prevention Program (NAEPP) guidelines are the prevailing clinical recommendation for children with asthma. The Expert Panel Reports presenting clinical practice duielines for the diagnosis and management of asthma have organized recommendations for asthma care around four components considered essential to effective asthma management: - Measures of assessment and monitoring, obstained by objective tests, physical examination, patient history and patient report, to diagnose and assess the characteristics and severity of asthma and to monitor whether asthma control is achieved and maintained. - Education for partnership in asthma care - Control of environmental factors and comorbid conditions that affect asthma - Pharmacologic therapy This section of the report updates information on each of these four components based on the Expert Panel's review of the scientific literature. The sections that follow present specific clinical recommendations for managing asthma long term and for managing exacerbations that incorporate the four compoenents. Two evidence tables were prepared: (1) Predictors of Exacerbation: https://www.nhlbi.nih.gov/health-pro/guidelines/current/asthma-guidelines/evid_tbls/1_predexacer.pdf (2) Usefulness of Peak Flow Measurement: https://www.nhlbi.nih.gov/health-pro/guidelines/current/asthma-guidelines/evid_tbls/2_peakflowmeasure.pdf
Provide all other grades and definitions from the evidence grading system	Methodology for report: Overall Methods Used To Develop This Report **Background** In June 2004, the Science Base Committee of the NAEPP recommended to the NAEPP CC that its clinical practice guidelines for the diagnosis and management of asthma be updated. In September, under the leadership of Dr. Barbara Alving, M.D. (Chair of the NAEPP CC, and Acting Director of the NHLBI), a panel of experts was selected to update the clinical practice guidelines by using a systematic review of the scientific evidence for the treatment of asthma and consideration of literature on implementing the guidelines.

In October 2004, the Expert Panel assembled for its first meeting. Using EPR-2 1997 and EPR-Update 2002 as the framework, the Expert Panel organized the literature searches and subsequent report around the four essential components of asthma care, namely: (1) assessment and monitoring, (2) patient education, (3) control of factors contributing to asthma severity, and (4) pharmacologic treatment. Subtopics were developed for each of these four broad categories.

The steps used to develop this report include: (1) completing a comprehensive search of the literature; (2) conducting an indepth review of relevant abstracts and articles; (3) preparing evidence tables to assess the weight of current evidence with respect to past recommendations and new and unresolved issues; (4) conducting thoughtful discussion and interpretation of findings; (5) ranking strength of evidence underlying the current recommendations that are made; (6) updating text, tables, figures, and references of the existing guidelines with new findings from the evidence review; (7) circulating a draft of the updated guidelines through several layers of external review, as well as posting it on the NHLBI website for review and comment by the public and the NAEPP CC, and (8) preparing a final-report based on consideration of comments raised in the review cycle.

Preparation Of Evidence Tables

Evidence tables were prepared for selected topics. It was not feasible to generate evidence tables for every topic in the guidelines. Furthermore, many topics did not have a sufficient body of evidence or a sufficient number of high-quality studies to warrant the preparation of a table.

The Panel decided to prepare evidence tables on those topics for which an evidence table would be particularly useful to assess the weight of the evidence-e.g., topics with numerous articles, conflicting evidence, or which addressed questions raised frequently by clinicians. Summary findings on topics without evidence tables, however, also are included in the updated guidelines text.

Evidence tables were prepared with the assistance of a methodologist who served as a consultant to the Expert Panel. Within their respective committees, Expert Panel members selected the topics and articles for evidence tables. The evidence tables included all articles that received a "yes" vote from both the primary and secondary reviewer during the systematic literature review process. The methodologist abstracted the articles to the tables, using a template developed by the Expert Panel. The Expert Panel subsequently reviewed and approved the final evidence tables. A total of 20 tables, comprising 316 articles are included in the current update (see figure 1-1). Evidence tables are posted on the NHLBI Web site.

Ranking The Evidence

The Expert Panel agreed to specify the level of evidence used to justify the recommendations being made. Panel members only included

ranking of evidence for recommendations they made based on the scientific literature in the current evidence review. They did not assign evidence rankings to recommendations pulled through from the EPR-2 1997 on topics that are still important to the diagnosis and management of asthma but for which there was little new published literature. These "pull through" recommendations are designated by EPR-2 1997 in parentheses following the first mention of the recommendation. For recommendations that have been either revised or further substantiated on the basis of the evidence review conducted for the EPR-3: Full Report 2007, the level of evidence is indicated in the text in parentheses following first mention of the recommendation. The system used to describe the level of evidence is as follows (Jadad et al. 2000):

- **Evidence Category A:** Randomized controlled trials (RCTs), rich body of data. Evidence is from end points of well-designed RCTs that provide a consistent pattern of findings in the population for which the recommendation is made. Category A requires substantial numbers of studies involving substantial numbers of participants.

- **Evidence Category B:** RCTs, limited body of data. Evidence is from end points of intervention studies that include only a limited number of patients, post hoc or subgroup analysis of RCTs, or meta-analysis of RCTs. In general, category B pertains when few randomized trials exist; they are small in size, they were undertaken in a population that differs from the target population of the recommendation, or the results are somewhat inconsistent.

- **Evidence Category C:** Nonrandomized trials and observational studies. Evidence is from outcomes of uncontrolled or nonrandomized trials or from observational studies.

- **Evidence Category D:** Panel consensus judgment. This category is used only in cases where the provision of some guidance was deemed valuable, but the clinical literature addressing the subject was insufficient to justify placement in one of the other categories. The Panel consensus is based on clinical experience or knowledge that does not meet the criteria for categories A through C.

In addition to specifying the level of evidence supporting a recommendation, the Expert Panel agreed to indicate the strength of the recommendation. When a certain clinical practice "is recommended," this indicates a strong recommendation by the panel. When a certain clinical practice "should, or may, be considered," this indicates that the recommendation is less strong. This distinction is an effort to address nuances of using evidence ranking systems. For example, a recommendation for which clinical RCT data are not available (e.g., conducting a medical history for symptoms suggestive of asthma) may still be strongly supported by the Panel. Furthermore, the range of evidence that qualifies a definition of "B" or "C" is wide, and

the Expert Panel considered this range and the potential implications of a recommendation as they decided how strongly the recommendation should be presented.

Panel Discussion

The first opportunity for discussion of findings occurred within the "topic teams." Teams then presented a summary of their findings during a conference call to all members of their respective committee. A full discussion ensued on each topic, and the committee arrived at a consensus position. Teams then presented their findings and the committee position to the full Expert Panel at an in-person meeting, thereby engaging all Panel members in critical analysis of the evidence and interpretation of the data.

A series of conference calls for each of the 10 committees as well as four in-person Expert Panel meetings (held in October 2004, April 2005, December 2005, and May 2006) were scheduled to facilitate discussion of findings and to dovetail with the three cycles of literature review that occurred over the 18-month period. Potential conflicts of interest were disclosed at the initial meeting.

Report Preparation

Development of the EPR-3: Full Report 2007 was an iterative process of interpreting the evidence, drafting summary statements, and reviewing comments from the various external reviews before completing the final report. In the summer and fall of 2005, the various topic teams, through conference calls and subsequent electronic mail, began drafting their assigned sections of the report. Members of the respective committees reviewed and revised team drafts, also by using conference calls and electronic mail. During the calls, votes were taken to ensure agreement with final conclusions and recommendations. During the December 2005 meeting, Panel members reviewed and discussed all committee drafts.

During the May 2006 meeting, the Panel conducted a thorough review and discussion of the report and reached consensus on the recommendations. For controversial topics, votes were taken to ensure that each individual's opinion was considered. In July, using conference calls and electronic mail, the Panel completed a draft of the EPR-3: Full Report 2007 for submission in July/August to a panel of expert consultants for their review and comments. In response to their comments, a revised draft of the EPR-3: Full Report 2007 was developed and circulated in November to the NAEPP Guidelines Implementation Panel (GIP) for their comment. This draft was also posted on the NHLBI Web site for public comment in February 2007. The Expert Panel considered 721 comments from 140 reviewers. Edits were made to the documents, as appropriate, before the full EPR-3: Full Report 2007 was finalized and published. The EPR-3: Full Report 2007 will be used to develop clinical practice guidelines and practice-based tools as well as educational materials for patients and the public.

References

1. EPR. Expert panel report: guidelines for the diagnosis and management of asthma (EPR 1991). NIH Publication No. 91-

	3642. Bethesda, MD: U.S. Department of Health and Human Services; National Institutes of Health; National Heart, Lung, and Blood Institute; National Asthma Education and Prevention Program, 1991. 2. EPR-2. Expert panel report 2: guidelines for the diagnosis and management of asthma (EPR-2 1997). NIH Publication No. 97-4051. Bethesda, MD: U.S. Department of Health and Human Services; National Institutes of Health; National Heart, Lung, and Blood Institute; National Asthma Education and Prevention Program, 1997. 3. EPR-Update 2002. Expert panel report: guidelines for the diagnosis and management of asthma. Update on selected topics 2002 (EPR-Update 2002). NIH Publication No. 02 5074. Bethesda, MD: U.S. Department of Health and Human Services; National Institutes of Health; National Heart, Lung, and Blood Institute; National Asthma Education and Prevention Program, June 2003. 4. Jadad AR, Moher M, Browman GP, Booker L, Sigouin C, Fuentes M, Stevens R. Systematic reviews and meta-analyses on treatment of asthma: critical evaluation. BMJ 2000;320(7234):537-40. 5. NHIS. National health interview survey (NHIS 2005). Hyattsville, MD: National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention, 2005. Available at http://www.cdc.gov/nchs/about/major/nhis/reports_2005.htm. Link to the evidence tables themselves: http://www.nhlbi.nih.gov/health-pro/guidelines/current/asthma-guidelines/evidence-tables
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	**Systematic Evidence Review Overview** **Inclusion/Exclusion Criteria** The literature review was conducted in three cycles over an 18-month period (September 2004 to March 2006). Search strategies for the literature review initially were designed to cast a wide net but later were refined by using publication type limits and additional terms to produce results that more closely matched the framework of topics and subtopics selected by the Expert Panel. The searches included human studies with abstracts that were published in English in peer reviewed medical journals in the MEDLINE database. Two timeframes were used for the searches, dependent on topic: January 1, 2001, through March 15, 2006, for pharmacotherapy (medications), peak flow monitoring, and written action plans, because these topics were recently reviewed in the EPR-Update 2002; and January 1, 1997, through March 15, 2006, for all other topics, because these topics were last reviewed in the EPR-2 1997. **Search Strategies**

	Panel members identified, with input from a librarian, key text words for each of the four components of care. A separate search strategy was developed for each of the four components and various key subtopics when deemed appropriate. The key text words and Medical Subject Headings (MeSH) terms that were used to develop each search string are found in an appendix posted on the NHLBI Web site. **Literature Review Process** The systematic review covered a wide range of topics. Although the overarching framework for the review was based on the four essential components of asthma care, multiple subtopics were associated with each component. To organize a review of such an expanse, the Panel was divided into 10 committees, with about 4-7 reviewers in each (all reviewers were assigned to 2 or more committees). Within each committee, teams of two ("topic teams") were assigned as leads to cover specific topics. A system of independent review and vote by each of the two team reviewers was used at each step of the literature review process to identify studies to include in the guidelines update. The initial step in the literature review process was to screen titles from the searches for relevancy in updating content of the guidelines, followed by reviews of abstracts of the relevant titles to identify those studies meriting full-text review based on relevance to the guidelines and study quality. The combined number of titles screened from cycles 1, 2, and 3 was 15,444. The number of abstracts and articles reviewed for all three cycles was 4,747. Of these, 2,863 were voted to the abstract Keep list following the abstract-review step. A database of these abstracts is posted on the NHLBI Web site. Of these abstracts, 2,122 were advanced for full-text review, which resulted in 1,654 articles serving as a bibliography of references used to update the guidelines, available on the NHLBI Web site. Articles were selected from this bibliography for evidence tables and/or citation in the text. In addition, articles reporting new and particularly relevant findings and published after March 2006 were identified by Panel members during the writing period (March 2006-December 2006) and by comments received from the public review in February 2007.
Estimates of benefit and consistency across studies	In summary, the NAEPP "Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma-Full Report 2007" represents the NAEPP's ongoing effort to keep recommendations for clinical practice up to date and based upon a systematic review of the best available scientific evidence by a Panel of experts, as well as peer review and critique by the collective expertise of external research/science consultants, the NAEPP CC members, guidelines implementation specialists, and public comment. The relationship between guidelines and clinical research is a dynamic one, and the NAEPP recognizes that the task of keeping guidelines' recommendations up to date is an increasing challenge. In 1991, many recommendations were based on expert opinion because there were only limited randomized clinical trials in adults, and almost none in children, that adequately tested clinical interventions grounded in research findings about the disease process in asthma. The large gaps in the literature defined pressing

	clinical research questions that have now been vigorously addressed by the scientific community, as the size of the literature reviewed for the current report attests. The NAEPP is grateful to all of the Expert Panel members for meeting the challenge with tremendous dedication and to Dr. William Busse for his outstanding leadership. The NAEPP would particularly like to acknowledge the contributions of Dr. Gail Shapiro, who served on NAEPP Expert Panels from 1991 until her death in August 2006. Dr. Shapiro provided valuable continuity to the Panel's deliberations while simultaneously offering a fresh perspective that was rooted in observations from her clinical practice and was supported and substantiated by her clinical research and indepth understanding of the literature. Dr. Shapiro had a passion for improving asthma care and an unwavering commitment to develop evidence-based recommendations that would also be practical. Dr. Shapiro inspired in others the essence of what NAEPP hopes to offer with this updated Expert Panel Report: a clear vision for clinicians and patients to work together to achieve asthma control.
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

Table 7

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• Interventions to Modify Health Care Provider Adherence to Asthma Guidelines: A Systematic Review • Sande O. Okelo, Arlene M. Butz, Ritu Sharma, Gregory B. Diette, Samantha I. Pitts, Tracy M. King, Shauna T. Linn, Manisha Reuben, Yohalakshmi Chelladurai and Karen A. Robinson. • September 2013 • Systematic Review Okelo et al, Pediatrics 2013 132:3:S17-34 • http://pediatrics.aappublications.org/content/pediatrics/early/2013/08/20/peds.2013-0779.full.pdf
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Demonstrates several tools are effective in enhancing the quality of care and reduce undesirable outcomes.
Grade assigned to the **evidence** associated with the	

recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	We followed the Agency for Healthcare Research and Quality Methods Guide for Effectiveness and Comparative Effectiveness Reviews (available at www. effectivehealth care.ahrq.gov/methods guide.cfm). Our protocol and the full report were subject to review. Data sources included Medline, Embase, Cochrane CENTRAL Register of Controlled Trials, Cumulative Index to Nursing and Allied Health Literature, Educational Resources Information Center, PsycINFO, and Research and Development Resource Base in Continuing Medical Education up to July 2012. Paired investigators independently assessed study eligibility. Investigators abstracted data sequentially and independently graded the evidence. RESULTS: Sixty-eight eligible studies were classified by intervention: decision support, organizational change, feedback and audit, clinical pharmacy support, education only, quality improvement/pay-forperformance, multicomponent, and information only. Half were randomized trials (n = 35). We identified 4217 unique citations of which 68 studies were eligible.
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

Table 8

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• Cochran Database of Systematic Reviews: Intermittent versus daily inhaled corticosteroids for persistent asthma in children and adults (Review) • Chauhan BF, Chartrand C, Ducharme FM • December 12, 2012 • Cochrane Database of Systematic ReviewsIntermittent versus daily inhaled corticosteroids forpersistent asthma in children and adults (Review)Chauhan BF, Chartrand C, Ducharme FMChauhan BF, Chartrand C, Ducharme FM.Intermittent versus daily inhaled corticosteroids for persistent asthma in children and adults.Cochrane Database of Systematic Reviews 2013, Issue 2. Art. No.: CD009611. • ~~Cochrane Review: Chauhan et al Cochrane Database Syst Rev 201212:CD009611~~ • https://www.ncbi.nlm.nih.gov/pubmed/23235678 • http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009611.pub3/epdf
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Different approaches to treatment achieve different outcomes in children and adults (Daily achieves better asthma control than intermittent inhaled corticosteroids)
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	**Search methods** We searched the Cochrane A irways Group Specialised Register of trials (CAGR) and the ClinicalTrials.gov web site up to October 2012. **Selection criteria** We included randomised controlled trials (RCTs) that compared intermittent ICS versus daily ICS in children and adults with persistent asthma. No co-interventions were permitted other than rescue relievers and oral corticosteroids used during exacerbations. **Data collection and analysis** Two review authors independently assessed trials for inclusion, meth odological quality and e xtracted data. The primary e fficacy outcome was th e number of patients with one or more exacerbations requiring oral corticosteroids and the pr imary safety outcome was th e number of patients with serious adverse health events. Secondary outcomes included exacerbations, lung function tests, asthma control, adverse effects, withdrawal rates and inflammatory markers. Equivalence was assumed if the risk ratio (RR) estimate and its 95% confidence interval (CI) were between 0.9 and 1.1. Quality of the evidence was assessed using GRADE.
Provide all other grades and definitions from the evidence grading system	We identified trials from the Cochrane Airways Group Specialised Register of trials (CAGR), which is derived from systematic searches of bibliographic databases including the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, EMBASE, CINAHL, AMED, and PsycINFO, and handsearching of respiratory journals and meeting abstracts (see Appendix 1 for further details). All records in the CAGR coded as 'asthma' were

	searched using the following te r ms: (intermittent* or as-needed* or “as needed” or pr n or irregular* or occasional* or sporadic* or short-course*) and (daily* or regular* or routine*). We also conducted an advanced search of ClinicalTrials.gov using ’intermittent’ as keyword, ’asthma’ as condition and ’interventional studies’ as study ty pe. All databases were searched from their inception to October 2012 and th ere was no restriction on language of publication. The search for literature conducted until to December 2011 identified a total of 206 citations and abstracts through database searching and 26 citations from clinicaltrials.gov . Of th em, 16 full-text potential trials were reviewed and finally six trials (seven comparisons) were included for the meta-analysis (Figure 1). We updated the literature search in October 2012. There were 6 additional references, but no new included studies. This review summarises the best evidence available up to October 2012 derived from six trials (1211 patients with suspected or confirmed persistent asthma) of high methodological quality. The results pertain to children and adults with persistent asthma, and preschoolers with repeated wheezing suspected of persistent asthma. The systematic search to identify eligible trials and unpublished reports minimise the risk of inclusion bias. The outstanding collaboration of the authors/funders of six of the seven comparisons (Martinez 2011a; Martinez 2011b; Papi 2007; Papi 2009; Turpeinen 2008; Zeiger 2011) allowed us to obtain additional unpublished data and confirmation of methodological quality which strengthened the meta-analysis. Due to th epaucity of trials or the absence of events, 11 of 37 secondary outcomes could not be aggregated. While study authors reported enrolling patients with confirmed or suspected persistent asthma, the criteria used in paediatric trials (frequency of exacerbations with or without atopy, family history of asthma and eosinophilia) may have included an unknown propor tion of preschool children with intermittent viral-induced asthma that may have diluted the effect. The review is heavily weighted towards preschool- and school-aged children, with only two trials pertaining to adults. The long-term impact of intermittent versus daily ICS on lung growth, airway remodelling, bone mineralisation and adrenal function in children and lung function decline in adults beyond one year of follow-up remain to be addressed. **Quali ty of the evidence** The included trials were of high methodology and were generally at low r isk of bias. The confirmation of methodology by almost all authors or funders (with supportive evidence such as study protocols) and the provision of additional unpublished data allowed more precise estimates. The quality of evidence for our key outcomes reflects a lack of power from the studies that we included in the analysis (statistical imprecision) and variation in the different

	approaches used (indirectness). No potential biases were found in the review process.
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

Table 9

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	• Quality of Care for Childhood Asthma: Estimtating Impact and Implications • Soeren Mattke, Francisco Martorell, Priya Sharma, Floyd Malveaux, Nicole Lurie • 2009 • S Mattke, et al. Quality of Care for Childhood Asthma: Estimating Impact and Implications. Pediatrics 123 Suppl 3, S199-S204. 3 2009. • https://www.ncbi.nlm.nih.gov/labs/articles/19221164/ • http://pediatrics.aappublications.org/content/pediatrics/123/Supplement_3/S199.full.pdf • ~~Systematic Review Mattke et al, Pediatrics 2009 123 S199-204~~
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	Identified multiple gaps in asthma care quality. Key outcomes identified include hospitalizations and emergency department visits. Identified large racial disparities in use of inhaled corticosteroids
Grade assigned to the **evidence** associated with	

the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	We conducted a review of 164 relevant publications to consolidate the evidence on gaps in the quality of asthma care, the impact of those gaps, and the costs and benefits of closing those gaps. To identify relevant publications, a comprehensive review was performed of the English-language literature dating from 1995 to 2006, using the key words asthma, quality of care, treatment, care, therapy, disparities, inadequate, variability, differential, inequity, gap, variation, variance, medication, adherence, utilization, guideline adherence, disease management, impact, effect, morbidity, mortality, hospitalization, hospital admission, emergency room, emergency department, loss, absence, work, employment, workdays, school, attendance, absenteeism, and presenteeism. To obtain abstracts of articles published in peer-reviewed journals, the traditional health literature databases (ie, Medline, Embase, CINAHL, PsycINFO, Social Sciences Abstracts, and EconLit) were searched. Reference lists in relevant articles were mined for additional items. In addition, a number of databases for non–peer-reviewed literature and the Web sites of relevant governmental, professional, and advocacy organizations were searched. Two reviewers (Drs Mattke and Martorell) independently reviewed the titles of identified items, to assess whether the articles were likely to provide information on gaps in asthma care and their impact. The 2 reviewers then independently reviewed the abstracts or, if abstracts were unavailable, the full publications, to determine whether the publications contained information relevant to our study. Differences between reviewers were resolved through consensus. Reports were omitted from further consideration if they did not contain any data relevant to our research questions, were not conducted in the United States, were review or opinion articles, or were duplicative publications of the same data.
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies	

conducted since the SR. Do the new studies change the conclusions from the SR?	

Overarching statement: Even when not specifically indicated, we are interested in how these constructs are impacted by such factors as race, ethnicity, socioeconomic status or its indicators, or the presence of other special health care needs.

Our metric is designed to capture axes related to two distinct conceptual frameworks:

1) Asthma is a model of chronic disease management. In other words, ED visits may arise from acute exacerbations indicating a flare up of disease, and/or suboptimal management of the chronic illness.
2) ED visits for asthma may reflect limitations of primary care beyond the provision of suboptimal treatment, such as insufficient education, limitations of access or availability, breakdowns of communication, or a variety of other factors.

We note that the internal quality of the ED visit to manage the asthma is not the target of this measure. However, communication between the emergency department and the primary care site may prove to be within the scope of this measure, pending the views of our experts and developers.

Construct I: Need to sufficiently specify population for measure

Concept	Implications (Lay Statement)	Lit Review Questions
(Descriptive) **The measure will need to adequately specify the population that we consider to be eligible for an ED with asthma measure.**	The development of measures regarding ED use for children with asthma requires us to understand the strengths and weaknesses for our measure of various approaches to identifying whether or not children have asthma. It further requires us to understand the impact of the availability of various sources of data (such as encounter data, pharmaceutical data, electronic medical record or chart review data) on these strengths and weaknesses. We are aware that the use of the term asthma is variable. We are not interested in diagnoses with the name asthma, but with an operational diagnosis that we will functionally treat as asthma, whether it has been called chronic wheezing, reactive airway disease, chronic infectious bronchitis, etc. We recognize that asthma and its presentation may change over the course of a child's life.	1. When asthma care is evaluated, how is the population of care recipients defined? How is asthma defined? What is the impact of including various types of data (dx 1 or more, drugs, etc) on the sensitivity and specificity of asthma identification? What are practical and valid approaches to identifying asthma? How do the answers to these questions differ between adults and children? 2. Are any groups persistently excluded from studies of asthma care (i.e., are children who have asthma and other comorbid conditions, such as a malignant disease, excluded?). What rationale is provided for the exclusion? 3. Are any non-asthma diagnoses considered to be indicators of asthma or potential asthma (e.g. bronchitis, bronchiolitis, wheezing, atopy) 4. For children up to age 21, how do issues of diagnosis, management, and follow-up differ by age and developmental stage? 5. At what point does literature suggest that reactive airway disease should be managed as asthma? a. What other conditions are managed as asthma? 6. What common current or preexisting comorbid conditions alter the management plan for asthma?

Construct II: Adequacy of management of asthma (as a chronic disease example)

Concept	Implications (Lay Statement)	Lit Review Questions

IIA. **↑Adequacy of asthma management:** **↓ED visits**	Since asthma is a chronic disease characterized by acute exacerbations, the extent to which asthma care is optimized through the use of appropriate medications, the control of the environment, and the preparation of the parent/child dyad to adapt to changes in circumstances (e.g. viral respiratory infection or exposure to cold) should reduce the number of ED visits, irrespective of the number of primary care visits.	1. What are the recommendations of the NHLBI guidelines? a. What does the literature suggest about the usefulness of NHLBI guidelines? b. Are there aspects that it has identified that appear to be missed? 2. What do we know about asthma management, how it's measured, who provides it, patterns of care and how ED visits vary as a consequence? 3. Does identification of PCP improve outcomes of ED visit, including patterns of care, utilization? 4. What do we know about the content of an asthma plan and its relationship to a full program of chronic disease management, and its influence on ED utilization? 5. What evidence is there about the impact on outcomes such as ED use when the child or adolescent is involved in asthma self-management? For example, does it matter if: a. The child has a written asthma plan? b. The child understands their asthma plan? c. The child is given an opportunity to participate in managing care? 6. How is the role of the child in self-management measured? 7. How much are children able to recognize, communicate and act on their asthma? 8. What do we know about the impact of asthma services on asthma management? This includes: a. Treatment from an asthma specialist; b. Social worker; or c. Multidisciplinary personnel 9. To what extent is ED use by children with asthma stimulated by non-asthma related issues? a. How can we identify when that occurs?

		b. What is the evidence that providing other services will reduce the number of ED visits? 10. To what extent do children contribute to their management (including avoiding triggers, recognizing symptoms, medication adherence, etc.)? a. What is the impact and variance by age? 11. What is the evidence regarding adequacy of various medication delivery systems for infants, toddlers, children and adolescents in acute and chronic settings? 12. Is there evidence of prior insult to the lungs such as sequelae of prematurity, etc. that create distinct subpopulations when considering this measure (at risk for ER visit)? 13. What aspects of the health services environment have been identified as contributing to outcomes of asthma management (e.g. school based health care)? 14. Does rate of ED utilization for non-respiratory diagnoses vary between asthmatics and non-asthmatics? 15. What is known about how often children with asthma use the ED over an extended period of time? Does it change over the life course of childhood? How does that vary by child characteristics, including race, SES, urban, suburban vs. rural, and age?

IIB. **↑PCP capacity/knowledge/skill:** **a. ↑Asthma management** **b. ↓Asthma exacerbations** **c. ↑Chronic disease management**	Broadly speaking, patient management of asthma is influenced by the capacity of the PCP practice. This includes the knowledge and skills possessed by the PCP, as well as office support to enhance access and availability of care. PCP includes the ability of the PC office to meet the cultural needs of the patient and their family.	1. What are the diversity of practices or services that may or may not impact ability or capacity of the PCP practice to manage asthma? 2. What do we know about the specific skills and processes that contribute to a primary care practice's capacity? 3. What patterns of visits or medication use or other indicators have been used as markers of well or poorly delivered primary care for asthma in children and/or adults? 4. What is the minimum use of specialists appropriate for children with asthma? How does that vary with history of ED or hospital use? a. When and how does the use of specialists become a marker for higher or lower quality of care? 5. What evidence is there regarding the nature of the PCP practice for children with asthma? For example, the level of continuity with individual clinicians vs. practices, the accessibility of specified clinicians and/or practices during the day and/or after work hours, etc.
IIC. **↑Asthma education:** **a. increases recognition of symptoms >** **b. ↑Management skills**	Enhancing what patients or their families know about asthma may be an important tool to improve care for children with asthma. The likely first effect of such education is to enhance the capacity of a caregiver to identify what symptoms may relate to asthma. This could conceivably increase utilization of both PCP and ED services if this were to increase the caregiver's perceived	1. What are metrics or processes regarding the quality of asthma care? Is it drug ratios (i.e. proportion of prescriptions filled that are for rescue vs control medications), asthma action plan, , capacity of PCP office, relationship to PCP practice, or other specific bundles of care, etc? 2. What constitutes "perfect care"/"best practice" for any specified type of patient? 3. What do we know about the impact of asthma education programs on quality of care, outcomes of care, or utilization of care? Define utilization of care as including: a. PCP utilization, b. ED utilization, c. Referral/specialist utilization,

	need for care for their child's asthma. With a more sophisticated understanding, including having a valid asthma action plan and understanding how to use it, ED care may be reduced and PCP care for asthma may be reduced, as symptoms are less frequent and parents are more competent to manage them when they arise.	d. Non physician care team member utilization, e. Medication usage, f. Hospitalizations, and/or g. Other care utilization areas to consider? Examples may include functional status, quality of life elements, spirometry, role functioning. 4. What is the diversity of asthma education programs and what are the differences in quality of care/outcomes/utilization of care associated with differences? 5. Does referral to an asthma specialist impact quality of care, utilization of care and asthma outcomes? 6. Does referral to a social worker impact utilization of care and asthma outcomes? 7. (Broad) Does involvement of multidisciplinary personnel (beyond allopathic or osteopathic physicians) impact quality of care, utilization of care and asthma outcomes? 8. What are desirable roles and effectiveness of interventions that extend beyond the healthcare system, such as reducing pollution, focusing on environmental justice, housing, dust mites, etc.? 9. How does organization and capacity of the practice setting influence the delivery of asthma management education?
Construct III: **Concept** **IIIA.** **↑Primary care capacity:** **a. ↑ PCP visits (routine, WCC)**	Adequacy of PCP practice site to Implications (Lay Statement) In general, enhanced capacity may affect a patient's access to	handle acute exacerbations of chronic disease and/or acute illnesses Lit Review Questions

b. ↑PCP visits (other acute dx) **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)** **IIIA.2** **SUBCONSTRUCT:** **↑Accessibility:** **a. ↑ PCP visits (routine, WCC)** **b. ↑PCP visits (other acute dx)** **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)**	care. Capacity can refer to patient services that make it easier for a patient to receive timely care, such as location or hours of offices, to the ability to triage phone calls in a timely and effective way, or may include the materials and services present within an office (e.g. the presence of a treatment room, the capacity to deliver oxygen, nebulizers, etc.) Such capacity may be limited or enhanced by staffing, space, the ability to safely transport someone from the office to a hospital, etc. If PCP office capacity is optimized, ED visits may be reduced as acute and mundane conditions can be managed in a PCP setting. Subsequently, increased capacity of the entire PCP support network will increase number of PCP visits.	1. What do we know about access to the PCP's office as a place to manage asthma, and the subsequent capacity of a PCP and the diversity of practice settings? Additionally, how do we measure capacity and, its impact on QoC, processes of care, asthma outcomes, asthma specific processes and utilization? How do these factors impact ED use or other outcomes? a. In general: i. PCP/specialist ratio in a plan or PCP/child ratio ii. PCP time spent in visit (incl. minutes per sick, well-child, asthma management visit) iii. Nature of training activities iv. How long does it take to schedule a visit (incl. asthma (chronic), acute, follow-up visit) v. Office hours and visit flexibility (incl. after hours coverage, office consult, meet in ED) vi. Phone capabilities: (incl. answering capacity, putting on hold, returning calls, after hours phone service) vii. Level of implementation of patient centered medical home/chronic care model, eg i. Use of registries ii. Standardized tools for measurement iii. Case management iv. Group visits or other education, etc

b. Specifically, ability to manage acute dx in office, which includes:
    i. Do they have a treatment room or capacity to use a room as a treatment room?
    ii. Do they offer rescue treatments (e.g. nebulizers, spacers)?
    iii. Can they measure oxygen saturation?
    iv. Do doctors feel comfortable with acute asthmatic in office?
    v. Can they take time to manage an acute pt in their office?
    vi. Do they have safe and rapid transport to a hospital (how long?)
2. Availability and accessibility of offices (incl. office hours, geographic distribution)
    a. What do we know about linguistic capabilities in the PCP setting influencing use of the ED?
    b. What do we know about proximity of the PCP office to public transit on the utilization of the ED?
3. What do we know about the impact of variations in patterns of care/practice, use of modalities, and/or and receipt of well-child care on asthma management or outcomes (eg ED use)? Does Immunization status reflect on t eh capacity of the PCP, on the state of the child, or on other factors that may relate to asthma outcomes? How about the sufficiency of the number of WCC Visits (eg meets HEDIS standard or AAP standard or does not)? Absolute number of visits to PCP?
4. Are children with more WCC visits less likely to use the ED for acute visits? children who are UTD on their immunizations?
5. What literature is there on the relationship between pediatric ED use

		and other measures of asthma exacerbation/outcomes? 6. What do we know about variability of capacity and management of mundane conditions (e.g. OM, URIs, pharyngitis), office to ED ratios? 7. What do we know about variability of capacity and management of acute conditions requiring interventions (e.g. asthma)? 8. To what extent does ED capacity increase use of ED services? Do hospitals advertise ED services, have fast track for mundane conditions, etc? 9. To what extent does ED have capacity to provide primary care, routine immunizations, etc? How is that built into policies and protocols? 10. At what age does the PCP start meeting alone with child? Time spent in visit? 11. To what extent and at what age do PCP's involve children in self-management and does it vary?

IIIB. **↑Relationship with PCP:** **a. ↑ PCP visits (routine, WCC)** **b. ↑PCP visits (other acute dx)** **c. ↑ PCP visits (asthma)** **d. ↓ED visits (acute dx, asthma)**	Improved relationship with PCP may increase visits to your PCP and decrease ED visits, for both acute and mundane conditions. A good relationship may lead to greater trust and adherence to recommendations (both WCC and asthma care) and drive a preference for seeking care by the PCP over seeking care in another environment. In general, we are referring to relationship of caregiver with PCP and their office staff. We recognize the importance of the relationship of PCP's with patients as well; when the relationship between the PCP and the child rather than caretaker is emphasized in research, we'd like to capture that as well.	1. What exists regarding measuring the quantity and quality of the relationship with PCP? Specifically: a. What's the variation and does it matter? b. How is it measured? c. What do we know about patient experience of care, especially as it relates to relationship with clinicians/PCP d. To what extent is quality of relationship expressed in terms of caregiver vs. child relationships and how does this change with age of child or longevity of connection to a PCP? 2. What evidence is there regarding use of supplemental services outside of regular clinical visits and how do these services impact quality and utilization of care? Define supplemental services as: a. Electronic educational/reminder tools (incl. social media) b. Telephone educational/reminder tools c. Print materials (e.g. educational brochures) d. Disease management, demand management, or other type programs e. Other services to consider? Measure quality, utilization of care should include at least : a. ED visits b. PCP visits 3. How does role of child in self care/management tie into these issues?

Construct IV: The connectedness of care in the primary care and ED setting – before, during, and after of the ED visit

Concept	Implications (Lay Statement)	Lit Review Questions

IV. (Descriptive) **Enhanced integration of ED care of asthma with routine care will have better outcomes**	If primary care is generally pretty good, then the ED visit should be an extraordinary event. In such cases the PCP alerting the ED to current management and the ED assuring appropriate follow up with the PCP is important. In cases where primary care is of lower quality or more variable, the ED visit may enhance the long term management of the child with asthma. And we need to assess this. One of the ways it might do so is to construct an asthma management plan that is then followed by the PCP. Another way is to connect a child without adequate primary care to primary care, especially to someone who is competent to manage the asthma.	1. What evidence supports that ED visits for asthma are most effective when visit is followed by a visit to the PCP? 2. Do utilization patterns in both the ED and primary care setting change following ED visits? 3. Is an effective/more effective use of medications seen following an ED visit? 4. Does the identification of a primary care provider improve outcomes of an ED visit (including patterns of care utilization)? 5. Is pre or intra visit communication with the primary care provider associated with better outcomes? How often does this occur? Are there systematic differences regarding those for whom this does and does not occur? 6. Are ED visits for asthma routinely associated with some form of communication or linkage with PCP? Does that result in better outcomes?

Construct V: Equity is a value in asthma care

Concept	Implications (Lay Statement)	Lit Review Questions
V. (Descriptive) **Equity is a critical construct of quality for children with equity**	Systematic differences in the frequency or nature of ED visits for asthma on the basis of race, ethnicity, family make-up, income/economic status, specifics of insurance status, presence or absence of comorbid special health care needs, etc represents decrements in quality that our measures should identify.	1. Does the literature indicate systematic or predictable differences in the frequency or nature of asthma care for children as it relates to ED visits for asthma that may be interpreted as representing inequitable structures, processes, outcomes, experiences with, or coordination of care? 2. What do we know about how social determinants and diagnosis and management of asthma and its outcomes, specifically as it relates to use of ED? 3. What do we know about the extent to which use of the ED for children with asthma that relates to the external physical and social environment?

Proposed Research Questions

Asthma- We propose to prioritize our Asthma Construct Table, to the following questions:

Acronyms

PCP: Primary Care Provider

ED: Emergency Department

WCC: Well-child care

Baseline Question (for Questions 1, 2 and 3 below):
When asthma care is evaluated, how is the population of asthma care recipients defined at the population level? What are specific implications of how you identify patients with asthma, including various approaches to specifying the denominator of children with asthma? What are practical and valid approaches to identifying asthma at the population level? How do the answers to these questions differ between adults and children?

Question 1 (Construct IIA.2):
For children with asthma, what do we know about asthma management? How is management of asthma described and measured? This includes who (PCP, asthma specialist, ED, etc) primarily manages it as well as who provides it. What are the patterns of care and what do we know about how use of the ED varies as a result of various approaches to management?

- **Question 1a (Construct IIB.3):**

Specifically, have any of these patterns of visits or medication use or other characteristics of care been used as markers of well or poorly delivered primary care for asthma for children and/or adults?

Question 2 (Construct IIB.5):
How has varying asthma care for children been described on the basis of characteristics of the PCP offices or practices? For example, are they characterized by the level of continuity between individual clinicians, the level of conntiuity with any provider in the practice, the accessibility of specified clinicians and/or practices during the day and/or after work hours, etc?

- **Question 2a (Construct IIIA.3):**

What do we know about the impact of variations in patterns of care/practice, use of treatment modalities, and/or receipt of well-child care on asthma management or outcomes (e.g. ED use)? How about the sufficiency of the number of WCC Visits (eg meets HEDIS standard or AAP standard or does not)? Absolute number of visits to PCP?

Question 3 (Construct IIC.7):
(Broad) Does involvement of multidisciplinary personnel (beyond allopathic or osteopathic physicians) impact quality of care, utilization of care and asthma outcomes both within context of a primary care practice or in other clinical settings?

- **Question 3a. (Construct IIIB.2):**

What evidence is there regarding use of supplemental services outside of regular clinical visits and how do these services impact quality and utilization of care?

The following poster describing this measure was submitted for peer review and accepted and presented at the Annual Research Meeting of AcademyHealth in 2014.

New Pediatric Quality Measures Program (PQMP) Measure of Emergency Department (ED) Use for Children with Asthma

LC Kleinman, LE Soloway, CJ Homer, A Vella, A Ting, N Massenburg, B Rabin Fastman, A Vachon, A Balbierz, M Manice, EA Barrow, L Pickering, E Shemesh *Mount Sinai Collaboration for Advancing Pediatric Quality Measures (CAPQuaM); Icahn School of Medicine at Mount Sinai; National Institute for Child Health Quality (NICHQ), Northeast Business Group on Health, and NY State Department of Health*

Icahn School of Medicine at Mount Sinai

OBJECTIVES

To describe a new asthma outcome measure developed for the federal Pediatric Quality Measures Program by the Collaboration for Advancing Pediatric Quality Measures (CAPQuaM), an AHRQ-CMS CHIPRA Center of Excellence.

To describe the approach CAPQuaM is using for measure development

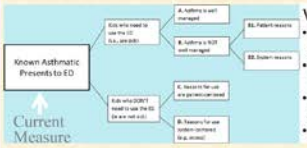
Consortium Partners include: AAP, AAFP, NICHQ, ACOG, CAHMI, NYS Medicaid, NCQA, Institute for Patient- and Family-Centered Care

CHIPRA = Child Health Insurance Program Reauthorization Act

METHODS

1. Qualitative Interviews with Clinicians/Patients
2. Literature Review
3. Criteria and Measure development
   - National multidisciplinary 9-person expert panel
   - 2 Round modified Delphi Process
   - Ratings of >250 clinical scenarios
     - •Inclusion/exclusion/reporting (59)
     - •What establishes ED as appropriate level of care (49)
     - •Establishing sufficiency of prior asthma care (59)
     - •Establishing root source of failures of prior management (34)
     - •Quality of ED Management (62)
   - Stakeholder Review and Input
4. Testing in NY State Medicaid Data

Overarching Conceptual Model



Who is a known asthmatic child?

- Prior asthma that health care plan should be able to identify
- Lower prevalence than asthma (~14-16% on Medicaid in NSCH)
- More prevalent than HEDIS persistent asthma (4.7% in NY State Medicaid)
- CAPQuaM Prevalence: 9.6%

Specifications: Assessing Eligibility and Scanning for Events Month by Month

Any prior hospitalization with asthma as primary or secondary diagnosis

Other Qualifying events after the fifth birthday (age is age at event):

One or more prior ambulatory visits with asthma as the primary diagnosis

(this criterion implies an asthma ED visit in the reporting month), **OR** Two or more ambulatory visits with asthma as a diagnosis, **OR**

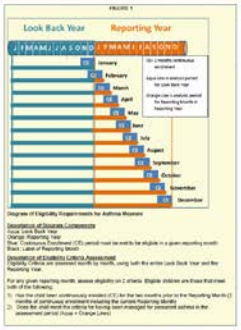
One ambulatory visit with asthma as a diagnosis **AND** at least one asthma related prescription, **OR**

Two or more ambulatory visits with a diagnosis of bronchitis

Other Qualifying events, any age:

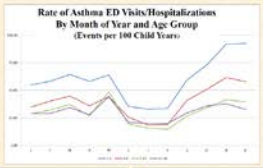
Three or more ambulatory visits with diagnosis of asthma or bronchitis, **OR** Two or more ambulatory visits with a diagnosis of asthma and/or bronchitis **AND** one or more asthma related prescriptions

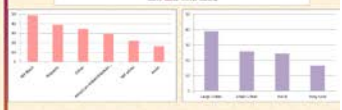
For eligibility purposes, asthma related medicine means long acting beta agonist (alone or in combination) or inhaled corticosteroid (alone or in combination), anti-asthmatic combinations, methylxanthines (alone or in combination), and/or mast cell stabilizers.



Numerator Events include Hospitalizations or ED Visits with Primary or Secondary Diagnosis of Asthma (most Medicaid systems would miss ED visits resulting in hospitalization if only ED visits sought)

Findings: NY Medicaid 2011





CONCLUSIONS

- CAPQuaM Asthma Measure is a true epidemiological rate of undesirable asthma outcomes (ED visits & hospitalizations)
- Measure performs well in NY State Medicaid data
- Captures variations, including racial

IMPLICATIONS

- CAPQuaM measures are now available for quality measurement and QI
- This measure improves upon currently available measures that assess undesirable utilization outcomes for children with asthma and should be adopted widely

FUNDING: AHRQ 1U18HS020518

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3219
Measure Title: Anticipatory Guidance and Parental Education
Measure Steward: Child and Adolescent Health Measurement Initiative
Brief Description of Measure: This measure is used to assess the degree to which pediatric clinicians discussed key recommended anticipatory guidance and parental education (AGPE) topics. Necessarily, anticipatory guidance questions vary by child age. Anticipatory guidance for children ages 0-9 months include 15 questions. Anticipatory guidance for children ages 10-18 months includes 16 questions; and anticipatory guidance for children ages 19-48 months includes 16 questions.
Developer Rationale: A primary component of well-child care is anticipatory guidance and parental education (AGPE). Past studies demonstrated that parents want to talk with health care providers about the topics that comprise anticipatory guidance and parent education recommendations. The Child and Adolescent Health Measurement Initiative (CAHMI) focus groups with parents and health care providers found that AGPE was the most important component of care provided in the context of discussions between the health care provider and the parent. Studies have shown that data derived from claims/billing codes and medical charts is not valid for determining whether specific topics were discussed and the degree to which the parent had their informational needs met on the specific topic. Parents are reliable and valid reporters of whether they recall discussions about specific topics and the degree to which their informational needs were met. The AGPE sections of the Promoting Health Development Survey (PHDS, see Attachment A-2. pages 8-10) focus on recommended topics for which there is evidence that providers can positively influence a parent behavior and only includes topics for which parents can reliably and validly report whether a discussion occurred. Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child´s health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews; they are not directly reported by parents.

Numerator Statement: The numerator is the number of parents who had a well child visit within the last 12 months and who indicated that they received anticipatory guidance and education, that their questions were answered or that they already had the information and did not require anticipatory guidance on that topic.
Denominator Statement: Parents whose children ages 0-48 months who received a well-child visit in the last 12 months and who responded to at least half of the AGPE items (see Attachment A-2 pages 8-10) on the Promoting Healthy Development Survey (PHDS: www.wellvisitsurvey.org)
Denominator Exclusions: Unknown and missing values (responses coded missing) are excluded in the data analysis. Approximately 2.6% of parents who started the Online PHDS did not complete the survey (range 0.0-3.3% for top 5 providers with highest number of surveys; see Testing form, pages 23-24 for more detailed information on missing data).

Measure Type: Outcome: PRO
Data Source: Other
Level of Analysis: Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary

- This is a Patient-Reported Outcome-Based Performance Measure (PRO-PM) derived from the responses to 23-25 questions (depending on age of child) on the Promoting Healthy Development Survey (complete survey starts on page 20 of the Appendix).
- The developer provided a logic model in both graphic and narrative: (1) the parent and child attend a well child visit with the provider; (2) the provider subsequently sends a survey -- the Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org), which includes one question (3 items, see Attachment A-2, page 17) for the parent to complete; (3) when at least 10 surveys have been completed, the provider receives a feedback report on parents' experiences of the visit and the extent to which they felt they received appropriate and adequate assessment of their family's alcohol use, substance abuse and safety via the CAHMI PHDS Toolkit website (www.phdstoolkit.org); (4) the provider reviews the report and then can engage in a Plan-Do-Study Act (PDSA) quality improvement process to improve his/her AGPE quality score.
- The developer also notes that its focus groups (with parents and providers) found "AGPE was the most important component of care provided in the context of discussions between the health care provider and the parent." The developer explains the "AGPE sections of the Promoting Health Development Survey focus on recommended topics for which there is evidence that providers can positively influence a parent behavior and only includes topics for which parents can reliably and validly report whether a discussion occurred". NQF staff review indicates the developer's citation indicates the topics are recommended by the American Academy of Pediatrics and Bright Futures.
- In the Performance Gap section the developer notes a HRSA study "found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to report their needs met for anticipatory guidance at the follow-up assessment than at the baseline assessment; and parents were more likely to be asked about one or more psychosocial (family assessment) topics at follow-up." The results are in included in the testing attachment.

Question for the Committee:

o *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Guidance from the Evidence Algorithm: Patient-reported outcome (Box 1)→ Relationship between PRO and provider action (Box 2)→ Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer reports that for the online PHDS, the proportion of parents who reported discussion of all anticipatory guidance and parental education topics or reported no need of discussion among unaddressed topics ranged 46.8-84.8% across the top 5 observed providers; all children averaged 60.0%.
- The results from the Kaiser Permanente Northwest Study (KPNW) study indicated 22.2% to 66.7% of children had parents reporting that providers provided the needed anticipatory guidance or parental education; the responses also varied by topic (injury prevention, physical care, child development and behavior). The proportion of all children meeting the criteria was 39.7%.

Disparities

- The online PHDS results showed variation according to a child's age (3-8 months=60.4%, 9-18 months=57.4%, 19-48 months=63.1%); race/ethnicity (Hispanic=58.9%, white=61.9%, black=56.9%, Asian=42.7%, other/multi-race=60.5%); level of risk for developmental, behavioral, or social delays (low/no risk=63.0%, high/moderate risk=56.2%) across all quality measures. For the Anticipatory Guidance and Parental Education measure, care for children 19-48 months, non-Hispanic white children, and children whose parents completed at least high school education are most likely to meet scoring criteria.
- After controlling for demographic and health factors, and provider differences, the KPNW study found differences by age, adjusted odds ratio of 0.55 (less than 9 months=63.1%, 10-18 months=55.8%, 19-49 months=47.4%).

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

- Evidence was presented that both parents and clinicians felt that anticipatory guidance was the most important point of discussion at well visits. Evidence was also presented to demonstrate that by using a survey tool followed by PDSA QI work, providers were able to achieve statistically significant improvement in providing effecting AGPE.
- This measure is an indicator of anticipatory guidance – a role the clinician needs to play in the child's development. Up to 16 different items that the clinician should discuss with the parent are asked for three different age groups. Two ways of creating the measure are discussed in this application. First, the measure is the percent of parents who responded "Yes" or "No but not needed" to all of the items. The second, which is the PHDS scoring algorithm, is the % of times the parent responded "Yes" or "No but not needed" to all the items and the clinician is provided the mean rate. We are being asked to consider the former. The measure is currently being used for improving care. The items are part of a larger survey entitled Promoting Health Development Survey. Parents complete the survey via web and when enough surveys are completed the physician gets various rates, including this one, and suggestions on how to improve them. This is an ongoing process so improvement can be measured. The items assess important development goals such as how the child communicates, reading to the child, the child's comprehension and specific parent needs like how to use a car seat and breastfeeding.
- This measure is a Process - Patient Reported Outcomes measure. The rationale is based in part on evidence from the Child and Adolescent Health Measurement Initiative focus groups with parents and health care providers finding that "AGPE was the most important component of care provided in the context of discussions between the health care provider and the parent." The measure proposes a method to determine how well providers are meeting the goal of providing anticipatory guidance and parental education to parents/care givers. Through post well-visit survey results, providers can learn whether they are effective in this goal and, if results are poor, training can improve effectiveness in this important part of care. The measured outcome is identified and supported by the rationale. Developer notes a HRSA study "found statistically significant and positive changes for the study interventions" (training sessions on AAP Bright Futures guidelines, based on the PHDS quality of care measures. Parents were more likely to report their needs for anticipatory guidance were met in follow-up assessments/surveys.
- There was a clearly demonstrated performance gap in providing AGPE with rates varying from only 40-60%. In addition, significant disparities among children by age, parental education and ethnic background were identified.
- The developers have provided three different well designed research studies that show substantial variation in the measure across socio demographic groups and that show there is room for improvement and real variability across providers.
- Yes, performance data on the measure was provided. A review of Promoting Healthy Development Survey (PHDS) results of the top 5 providers, a range of 46.8-84.8% was seen (proportion of parents who reported

discussion of all AGPE topics or reported no need of discussion among unaddressed topics). The cited KPNW study indicated a range of 22.2-66.7%. Disparities were shown with some groups achieving high scores.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Other – patient/family reported survey

Specifications:

- Level of analysis: Clinician – individual
- Interpretation of score: Better quality = Higher score
- This is a patient-reported outcome-based performance measure (PRO-PM)
- Numerator: *The numerator is the number of parents who had a well child visit within the last 12 months and who indicated that they received anticipatory guidance and education, that their questions were answered or that they already had the information and did not require anticipatory guidance on that topic.*
- Denominator: *Parents whose children ages 0-48 months who received a well-child visit in the last 12 months and who responded to at least half of the AGPE items (see Attachment A-2 pages 8-10) on the Promoting Healthy Development Survey (PHDS: www.wellvisitsurvey.org)* [Questions start on page 3 of Appendix A.]
- Exclusions: The developer states that "Unknown and missing values (responses coded missing) are excluded in the denominator for the data analysis. If a parent answered less than half of the items, those data are considered to be missing and were excluded from analysis." [NQF does not consider this an exclusion as it is defining the population of the measure.]
- The developer includes a calculation algorithm.
- The measure is not risk adjusted or risk stratified, but the developer states that it can be stratified by variables such as child demographics characteristics (e.g., the child´s age, race); child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or parent health characteristics, if large enough data sets are available.
- The measure does not use sampling.
- This measure relies on a set of 23-25 questions within the Promoting Healthy Development Survey (page 3 of the Appendix); the number and content of questions differs by age group (3-9 months, 10-18 months, 19-48 months). The questions are on three topics: physical care, developmental and behavioral guidance, and injury prevention. This online survey is initiated by the provider who sends it to a parent after a well-child visit. Providers must have a minimum of 10 surveys to generate a report to maintain parent confidentiality.

Questions for the Committee:

- *Are all the data elements (question items) clearly defined?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

NQF Note: Both measure score and data element reliability testing are required for PRO-PMs.

Method(s) of reliability testing

- The developer used data from the online Promoting Healthy Development Survey (PHDS), a KPNW study, and a HRSA evaluation study that tested "three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young children in the context of discussions between pediatric clinicians and parents during well-child visits"; the HRSA study used 5 tools, including the PHDS.
- The Cronbach alpha to test internal consistency (data/item element reliability) was calculated using the Online PHDS and KPNW data. Because each of the three age ranges has different questions, testing was done for each age group. In addition, factor analysis was performed to investigate the dimensionality of the scale.
- Score-level reliability was assessed by ANOVA and then intra-class correlation (ICC). Providers with 20 or more surveys were assessed; no information on the N is provided for these analyses.

Results of reliability testing

- Using the top 5 individual providers with the highest number of surveys (N=77 to 94) from the online PHDS testing, the developer reports the Cronbach's alphas for internal consistency (item-level) all fall within the acceptable range of 0.70-0.95. Scores for the 3-9 month group ranged from 0.76-0.98; for the 10-18 month group 0.85-0.94; and for the 19-48 month group 0.71-0.98. For all providers, scores ranged 0.89-0.90. Alphas were not available by provider for the topic areas, but for all providers, the physical questions scored 0.77, for behavioral/developmental=0.86, and injury prevention=0.73.
- Using the top 5 individual providers in the KPNW study, the developer reports the Cronbach's alphas for internal consistency (item-level) generally fall within the acceptable range of 0.70-0.95. Scores for the 3-9 month group ranged from 0.83-0.97; for the 10-18 month group 0.85-0.93; and for the 19-48 month group 0.83-0.88. For all providers, scores ranged 0.89-0.91. Alphas were not available by provider for the topic areas, but for all providers, the physical questions scored 0.71-0.80, for behavioral/developmental=0.80-0.85, and injury prevention=0.65-0.71.
- The developer reports that the intraclass correlation coefficient for the AGPE measure is 0.72, which according to the developer, citing to the literature, is "good" and in the acceptable range of 0.70-0.80.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance among clinicians can be identified?*

Guidance from the Reliability: Precise specifications (Box 1)→ Empirical reliability testing (Box 2) → Score level testing (Box 4) → Appropriate method used (Box 5) → High certainty or confidence that the performance measure scores are reliable (Box 6a)→ Moderate

Highest possible rating is HIGH.

Note: PRO-PMS *require* element-level testing as well, which was conducted. If judged without score-level testing, the highest eligible rating for this type of testing is MODERATE.

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

NQF Note: Both measure score and data element validity testing are required for PRO-PMs.

Method of validity testing of the measure score:

- ☒ **Face validity**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- Factor analysis was conducted to assess the construct validity of the quality measure. Pearson correlation coefficients were calculated between age-specific anticipatory guidance scales to assess the degree to which each of the item provide unique information.
- To assess the concurrent validity of the quality measure, hypothesized associations among PHDS items and scales were examined. The developer tested three hypotheses:
  "Respondents who indicate that providers talked with them about recommended anticipatory guidance topics are:
  - more likely to report increased confidence as a parent because of interactions with health care providers
  - more likely to report positive parenting behaviors in related areas
  - less likely to report being concerned about their child's development in related areas compared with respondents who indicate that providers did not talk with them although they wished they had done so."
- Pearson correlation coefficients were calculated across all scale measures to test hypotheses about expected relationships among the PHDS quality measures and to assess the degree to which each of the PHDS quality measures provide unique information. The developer notes that "We expect a moderate or strong correlation between the family assessment scale measures (>0.30) and inter-scale correlation coefficients to be less than 0.80."
- The PHDS survey also was tested using focus groups, in-depth cognitive interviews, a literature review, and an advisory board of expert stakeholders.

Validity testing results:

The developer reports the following results:

- Average factor loading for AGPE was 0.72; the developer states the acceptable level is above 0.60. Factor analysis suggests that the scale items are unidimensional.
- The concurrent validity testing results showed that parents reporting positive parenting behaviors had higher scores on the anticipatory guidance quality measure compared with parents not reporting positive behaviors:
  - "Parents who reported that their questions on specific anticipatory guidance topics were answered were more likely to report higher confidence in related parenting activities because of information and counseling received from their child's doctor or other health care providers compared with parents answering "no, but I wish we had discussed that" (odds ratio [OR]: 5.9, 95% confidence interval [CI]: 3.4-10.2; OR: 8.3, 95% CI: 5-13.8). Moreover, parents who reported positive parenting behaviors in the areas of injury prevention (70.9 vs 92.1, P < .000) and reading to their child (69.1 vs 13, P < .000) also had significantly higher scores on the "anticipatory guidance from providers" quality measure. In addition, significantly fewer parents reported concerns about their child's behavior if they also reported

that their child's doctor or other health care providers talked with them about the kinds of behaviors they might expect to see in their child (46.7% "yes, talked" vs 65.5% "no, wish", P < .000; OR: 0.46 95% CI: 0.29 – 0.72)."

- The developer provides a table of Pearson Correlation Coefficients, which assesses whether the measures are examining different topics. The results suggest, according to the developer, that the measures are not redundant, with an average correlation of 0.34. This measure was most highly associated with the *Family centered care measure* (0.52).

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:
N/A

Questions for the Committee:

- *Is the lack of exclusions consistent with the evidence?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- The developer does not risk adjust the measure because "we do not expect variation in the quality of care provided for children due to risk factors, e.g. children with special health care needs. The provider's performance should be the same regardless of risk factors."
- The developer notes the measure can be stratified by several demographic or health variables as "Identification of variation in quality measures across subgroups of children helps to highlight aspects of care and population of children for which preventive and developmental services may be most need of improvement."
- The developer reports that many studies have shown differences in access to and quality of care, as well as parent satisfaction. The developer states that "One study found: Overall, 94.0% of parents reported 1 or more unmet needs for a number of aspects of care, including assessing family alcohol use, substance abuse and safety. Uninsured children and children aged 18 to 35 months are disproportionately represented among the 15.3% of children whose parents indicated an unmet need this area of care. There are significant variations in performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors."
- Variations were observed by demographic and socioeconomic factors.

Question for the Committee:

- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer assesses meaningful differences among providers for the top 5 providers (number of individual surveys completed) in the online PHDS; across 56 providers using KPNW study data; and pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.

- Online PHDS: For the top 5 individual providers with the highest numbers of surveys (n=77 to 94), a range of 47.3%-84.8% of parents of young children reported that their needs for discussion on all items were met; the average for all children was 60.0%.
- KPNW Study: 39.7% of children had parents reporting that all of their discussion and education needs were met. Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant, 22.2% to 66.7% (p = 0.03). Provider n ranged from 15-153.
    - When the provider was used as the level 2 clustering variable, only 1.1-2.2% of the total variance observed was explained by either measured or unmeasured differences between providers. The developer indicates that this "suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers" and that "the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across the all quality measures" (i.e., providers are inconsistent and going to a different provider may not improve a child's care). However, the HRSA study does demonstrate that providers can improve their performance with an intervention.

Question for the Committee:

- *Does this measure identify meaningful differences in quality?*

2b6. Comparability of data sources/methods:
N/A

2b7. Missing Data
The developer reports the following:

- **Online PHDS:** Rate of survey completion was calculated based on survey start and complete dates for each respondent. According to the quality measure scoring protocol, if a parent answered less than half of the items in the AGPE measure, his/her score is considered to be missing. This does not include items that should have been appropriately skipped. Missing responses are not given a valid score and are not included in the calculation of the quality measure.
- **Online PHDS** data show that 2.6% of parents who started the survey did not complete the survey.
- For the online PHDS, overall missing data on this question was 1.9%, ranging from 0-3.3% for the top 5 providers.
- **KPNW Study:** Of the 5,755 sampled children, 2,173 surveys were returned (37.8%). For these children, the provider the parent identified and the provider to which the child was assigned by the health plan were the same 97.3% of the time. A 95% response rate was obtained for the provider survey.
- The developer notes that responses for the KPNW survey did not differ by gender or insurance type, but did differ by age and by number of previous well visits.
- The specifications indicate that **"**Unknown and missing values (responses coded missing) are excluded in the denominator for the data analysis. If a parent answered less than half of the items, those data are considered to be missing and were excluded from analysis."
- The developer states information about non-respondents is not available, but "Overall, the quality measure had less than 2% of missing cases, ranging 0-3.3% across the top 5 providers with highest number of surveys. Few overall missing values suggest that the measure level results unlikely to be biased by non-response to the survey questions."

Guidance from the Validity Algorithm: Specifications consistent with evidence (Box 1)→ Threats to validity addressed (Box 2) →Empirical validity testing (Box 3)→ Measure score testing (Box 6)→ Appropriate method (Box 7)→ Moderate certainty or confidence that the performance measure scores are a valid indicator of quality (Box 8b)→ Moderate

The highest possible score is MODERATE.

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
RATIONALE: Missing data not adequately addressed; non-respondent bias not available.

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

- All of the data elements are clearly defined, including specific questions, online survey tool. Some concern exists regarding parents' willingness to complete multiple surveys over time and how this will affect results.
- The data elements are clearly defined as they are part of an existing survey that has already undergone rigorous reliability testing at the item level. The algorithm is clear though it was challenging to figure out at first as it does not align with the way the survey is scored. Evidence has been provided that the survey can be consistently implemented but suffers from the usual bias associated with online vis paper forms.
- Reliability specifications were well defined.
- Reliability testing across providers was "good" between 70-80%. The number of surveys (>50) seems to be adequate.
- Reliability testing was done based on three different studies with adequate sample sizes and appropriate variability across socioeconomic status and age. Cronbach's alphas were consistently in the 80-90% range for all items together and in the 70-80% range for the items in each of the three questions. Though included as part of validity there was high factor loading on similar items and acceptable ICCs. With three studies the developers provide strong evidence for generalizability and sufficient evidence that performance among clinicians can be identified. While not included, it would have been beneficial to do some test-retest reliability assessments to see if parents are responding consistently.
- Reliability testing included three methods: review of data from the PHDS, a KPNW study and a HRSA evaluation. Data from the top 5 providers from the online PHDS testing was used in much of this assessment. Cronbach's alphas for internal consistency, used in review of PHDS testing and KPNW study - all fall within acceptable range.
- There was significant correlation between positive parenting scores and AGPE quality scores, which suggests this is a strong indicator of quality.
- Item testing was done during the build of the survey. Cognitive testing was done on each item as well as interviews to fine tune the questions. Factor analysis showed that the items loaded into one domain but ICC showed that there was enough variability across items to warrant each to be included. Concurrent validity was demonstrated between this measure and positive parenting behaviors. Discriminant variability was demonstrated by the low to moderate correlation between the Anticipatory Guidance domain and the other PHDC domains
- Validity testing for face validity and empirical validity was conducted.
- It is not clear how the missing data (for those parents who completed less than 50%) of the survey would affect validity of the results.
- Given the nature and use of the measure as an improvement indicator, risk adjustment is not needed. With results from the three studies we see substantial variability across providers and for children within providers. Given the concurrent validity results the measure does identify meaningful differences. Missing data is around 2% which is acceptable and an assessment showed little differences between those who completed the survey and those who did not.
- No exclusions were made as all patients/caregivers were to be included in the measure. This was also the case for risk adjustments as the quality of care should be consistent for all patients. Under Meaningful Differences was evaluated showing ranges of 47.3-84.8%. "Parents of young children reported that their needs for discussion on all items were met" - average for all children was 60.0%. Missing data, was assessed to to be not adequately addressed in preliminary review. Non-respondent bias also not available.

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- This is a patient reported outcome. Data are generated by parents completing the CAHMI-developed Promoting Healthy Development Survey (PHDS), which is sent to them by their provider following a well child visit.
- Although the survey has been in use since 2001, there is not currently an automated reporting system for providers. The developer has been working on a new website for the survey that will automatically report data, and expects it to launch in February 2017.
- The developer reports that the provider registration takes about 10 minutes and the parent survey takes about 15-20 minutes. There are no fees, licensing requirements, etc., to use the measure.

Questions for the Committee:

- *Is the data collection strategy ready to be put into operational use?*

o *Does the developer have a status update on the new website?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

- This measure does require provider registration with the survey tool and requires a patient reported outcome in the form of a 15-20 minute survey. Because of the extra effort required (mainly from the parental end), the reporting may not be as robust as desired.
- In terms of administration, the survey is already in use and has proven to be easily collected. That said, there was a substantial number of issues with the CAHMI website including times when it was unusable, it had issues with generating reports and tracking provider performance, CHAMI is working on a new website but there is no guarantee that the problems will be resolved. So, feasibility is dependent on an unknown at this point.
- In this patient reported outcome, data is obtained by parents completing the online PHDS, sent to them by the provider after a well child visit. At the point of submission of the measure, an automated reporting system for providers was still in development. It appears that this will need to be completed prior to operational use. The survey itself requires little time for the provider to register. The parent survey takes 10-20 minutes online. This should be feasible.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No** ☐ **UNCLEAR**

OR

Planned use in an accountability program? ☐ **Yes** ☒ **No**

Accountability program details

No confirmed use for an accountability program, but the developer has been in discussion with a number of organizations that are interested in using the measure, including CMS/Medicaid, Title V, and Head Start.

Improvement results

The developer provided the following response: "In a 2010-2012 study of a large pediatric practice in Oregon (n=551 providers), anticipatory guidance and parental education for physical care increased from 70.3% (n=379) at baseline (2010) to 77.6 (n=197, 2011-12, AOR:1.67, CI:1.11-2.50) post implementation of the CAHMI Well Visit Planner - a family engagement tool to assist parents in planning for their well child visit. This represents a 10.4% increase and was statistically significant at the 95% confidence level. The PHDS, which contains the Anticipatory Guidance and Parent Education measure, was used as the evaluation tool."

Unexpected findings (positive or negative) during implementation

The developer was not aware of any unintended consequences.

Potential harms
The developer was unaware of any potential harms.

Vetting of the measure
N/A

Feedback:
N/A

***Questions for the Committee**:*
- o *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

- This measure is not currently in use, but clearly achieving improvement in positive parenting would exemplify high quality healthcare.
- CAHMI is working on a new website but there is no guarantee that the problems will be resolved. So, feasibility is dependent on an unknown at this point. The survey has been in use for several years and in the past has been used as a tool for clinicians to improve their communication with parents. The proposed measure has not been used as yet but there is interest by CMS/Medicaid, Title V and Head Start to use this as a measurement indicator. The measure has the potential to be used to assess improved efficiencies in care and unintended consequences are really around deficiencies in the website.
- This measure is not being publicly reported. No unintended consequences are found. The use of this measure should support the provision of high quality anticipatory guidance and parent education in pediatrics.

Criterion 5: Related and Competing Measures

This measure is part of a set of five based on the PHD survey.
- 3219: Anticipatory Guidance and Parental Education
- 3220: Ask About Parental Concerns
- 3221: Family Centered Care
- 3222: Assessment of Family Alcohol Use, Substance Abuse and Safety
- 3223: Assessment of Family Psychosocial Screening

Harmonization
- o N/A

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as

demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

RATIONALE IF NOT ELIGIBLE: The measure has not been vetted.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 3219
Measure Title: Anticipatory Guidance and Parental Education (AGPE)
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: Click here to enter a date

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome
☒ Health outcome: Click here to name the health outcome
☒Patient-reported outcome (PRO): Anticipatory Guidance and Parental Education (AGPE)

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Figure 1 (attached) shows the logic model by which the anticipatory guidance and parental education (AGPE) quality measure is obtained and improved. Simply said: (1) the parent and child attend a well child visit with their provider; (2) the provider subsequently sends a survey -- the Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org), which includes three questions (15-16 items depending on the child's age) on AGPE (see Attachment A-2, pages 8-10) for the parent to complete; (3) when at least ten surveys have been completed, the provider receives a feedback report on parents' experiences of the visit and the extent to which they felt their anticipatory guidance and educational needs were met via the CAHMI PHDS Toolkit website (www.phdstoolkit.org); (4) the provider reviews the report and then can engage in a *Plan-Do-Study Act* (PDSA) quality improvement process to improve their AGPE score. THE PDSA cycle involves reviewing the baseline data; developing and implementing a plan of action to improve the score; obtaining further data from the parent; and comparing the first set of results with the second. The full process is repeated until providers are satisfied with their improved scores. We are currently applying for this process to be approved by the American Board of Pediatrics (ABP) for maintenance of certification (MOC, Part 4) credit. The provider must complete three PDSA cycles. Each time point must have at least 25 completed surveys and there must be at least 8 weeks between time periods.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

A primary component of well-child care is anticipatory guidance and parental education (AGPE). Past studies demonstrated that parents want to talk with health care providers about the topics that comprise anticipatory guidance and parent education recommendations. The Child and Adolescent Health Measurement Initiative (CAHMI) focus groups with parents and health care providers found that AGPE was the most important component of care provided in the context of discussions between the health care provider and the parent. Studies have shown that data derived from claims/billing codes and medical charts is not valid for determining whether specific topics were discussed and the degree to which the parent had their informational needs met on the specific topic. Parents are reliable and valid reporters of whether they recall discussions about specific topics and the degree to which their informational needs were met. The AGPE sections of the Promoting Health Development Survey (PHDS – see Attachment A-2, pages 8-10) focus on recommended topics for which there is evidence that providers can positively influence a parent behavior and only includes topics for which parents can reliably and validly report whether a discussion occurred.[1] Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child's health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews; they are not directly reported by parents. The process outlined in the logic model allows health care providers to better understand the extent to which

their patients experience "quality care" – in this case, the extent to which parents felt their anticipatory guidance and educational needs were met. It also allows providers to engage in quality improvement activities to improve their parent-reported AGPE quality scores by using several Plan-Do-Study Act (PDSA) cycles, as described above.

[1]Bethell, C, Rueland C, Halfon N and Schor, E. Measuring the quality of preventive and developmental services for young children: national estimates and patterns of clinicians' performance. Pediatrics, 2004.; 113; 1973.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	

What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

_____

1a.4 OTHER SOURCE OF EVIDENCE

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

Figure_1_AGPE_Logic_Model.docx,CAHMI_AGPE_evidence_attachment_revised_02_02_17.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

No

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

IF a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)

IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.

A primary component of well-child care is anticipatory guidance and parental education (AGPE). Past studies demonstrated that parents want to talk with health care providers about the topics that comprise anticipatory guidance and parent education recommendations. The Child and Adolescent Health Measurement Initiative (CAHMI) focus groups with parents and health care providers found that AGPE was the most important component of care provided in the context of discussions between the health care provider and the parent. Studies have shown that data derived from claims/billing codes and medical charts is not valid for determining whether specific topics were discussed and the degree to which the parent had their informational needs met on the specific topic. Parents are reliable and valid reporters of whether they recall discussions about specific topics and the degree to which their informational needs were met. The AGPE sections of the Promoting Health Development Survey (PHDS, see Attachment A-2. pages 8-10) focus on recommended topics for which there is evidence that providers can positively influence a parent behavior and only includes topics for which parents can reliably and validly report whether a discussion occurred. Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child´s health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews; they are not directly reported by parents.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis**. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

See also Testing Attachment_Anticipatory Guidance

DATA SOURCES:

Differences in the AGPE quality measure scores across providers is demonstrated for (1) 5 top individual providers with the highest number of surveys using Online PHDS data; (2) across 56 providers using Kaiser Permanente NW study data; and (3) pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.

Online PHDS: The performance scale for the AGPE quality measure was calculated using the scoring methods described in Attachment A-4 . Individual provider level differences in performance were illustrated by the proportion of children meeting the quality of care criteria across 5 top providers with the highest number of completed surveys after their well-child visit.

KPNW Study: The significance of differences observed in the proportion of children meeting criteria for the AGPE quality measure across pediatric providers (n=56) was evaluated using t-tests. The relative spread in the AGPE quality measure score across providers was assessed using the coefficient of variation statistics (standard deviation across providers multiplied by 100%). Multi-level regression models were conducted using the pediatric provider as the level 2 clustering variable in order to assess the degree to which the probability that a child meets criteria on each quality measure is explained by differences between providers (called the "clustering effect"). In implementing this multi-level regression method (Empty Model), the presence of a significant clustering effect by pediatric providers was estimated prior to accounting for the child and family characteristics associated with each provider. Second, variables related to the child and family characteristics (child's age, gender, race/ethnicity, birth order, developmental and behavioral delay risk status; parent education and risk for depression) were added to the Empty Model to assess how much of the provider clustering effect observed remains after accounting for these characteristics (called the "Patient Model").

HRSA study: Quantitative data results for the baseline (2010) and follow-up (2011-12) study of the intervention sites using the HRSA Evaluation Study data were conducted using basic descriptive statistics to describe each sample and applying chi-square test of statistical significance to assess differences in the quality measure for the baseline and follow-up samples.

PERFORMANCE RESULTS

Online PHDS: Table 1b.2a present the proportion of children whose care met for the AGPE quality measure across 5 providers. The proportion of parents who reported discussion of all anticipatory guidance and parental education topics or reported no need of discussion among unaddressed topics ranged 46.8-84.8% across 5 observed providers.

Table 1b.2a: Proportion of Children Meeting Measure Criteria, Top 5 individual providers with highest number of surveys

Characteristics (n=5355)	All Children					
	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)					
	1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)	
Children whose parents had their needs met on all items	60.0%	69.6%	47.3%	66.7%	84.8%	46.8%

KPNW Study: The proportion of children who had parents who had their anticipatory guidance and parental education informational needs met (meaning they reported either "yes, topic was discussed" or "no, but I already had information and did not need to talk about it" on each topic) was 58.8% on injury prevention topics (17.6% reported "yes, topic was discussed" to all topics), 56.2% on physical care topics (10.4% reported "yes" to all topics) and 52.7% on child development and behavior topics (13.1% reported "yes" to all topics). About two in five of children (39.7%) had parents who reported having their needs met across all topics. Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant (p=0.003).

Table 1b.2b: Proportion of all children in the study who met criteria for receiving quality services and ranges in proportion across providers.

Developmental Services Quality Measures All Children (n = 2173)	Proportion of Meeting Measure Criteria	Range in the Proportion of Children Meeting Measure Criteria Across 51 Pediatric Providers	Relative Variation (COV) in Measure Scores Across Pediatric Providers
Children whose parents had their needs met on all items	39.7%	22.2% to 66.7%	SD:10 (p = 0.03) 25.1%

Only providers with n=15 or more responses are included in the provider level analysis. Provider level sample sizes range from 15 to 153.

Multi-level analysis: For the Empty Model that used the provider as the level 2 clustering variable, only 1.1% to 2.2% of the total variance observed in whether children met criteria for all quality measures was explained by either measured or unmeasured differences between the providers that they see. This suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers. These findings translate into a 1.19 to 1.29 median odds ratio across the all quality measures in the PHDS, including AGPE, indicating that the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across all quality measures. When child/family level characteristics are added to the model (Patient Model), the total variance explained by differences between providers does not change significantly.

HRSA study

The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to report their needs met for anticipatory guidance at the follow-up assessment than at the baseline assessment; and parents were more likely to be asked about one or more psychosocial (family assessment) topics at follow-up. The tables below present comparison of percent of children who received care met the quality care criteria between baseline and follow-up survey data for each measure and overall composite comprehensive care measure.

Table 1b.2c: Anticipatory Guidance & Parent Education Measure by Children's Characteristics

Parent had their needs met on all AGPE topics

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	38.9% (216)	45.2% (146)	0.08
10-18 months	48.5% (208)	45.7% (150)	0.46
19-48 months	55.0% (193)	65.9% (147)	0.01
Race			
Hispanic	46.0% (46)	47.8% (46)	0.86
White	46.2% (475)	51.9% (372)	0.02
Asian	35.7% (10)	52.9% (9)	0.35
Multiple or other	62.5% (15)	33.3% (6)	0.12
Insurance type			
Private or private and public	46.4% (502)	49.9% (339)	0.15
Public only (includes Medicaid, Medicare, CHIP, and Military)	44.7% (85)	54.9% (89)	0.07
Other insurance type	(3)	(1)	-
Uninsured	50.0% (6)	(4)	-
At risk of developmental delay			
Low/no risk	47.5% (487)	52.1% (285)	0.09
High/moderate risk	40.7% (114)	44.4% (76)	0.49

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

NA

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

See also Testing Attachment Anticipatory Guidance

DATA SOURCES

We used the following data sources for testing of the AGPE quality measure:

(1) Online Promoting Healthy Development Survey (PHDS) – data collected through an online, publicly available tool (Promoting Healthy Development Survey-PHDS). Parents who had a well-child care visit in the last 12 months can complete the PHDS. Providers initiate the survey. (See Evidence Form Figure 1 for visual model of the Online PHDS.)

2) Kaiser Permanente Northwest (KPNW) Study – CAHMI partnered with Kaiser Permanente Northwest in Portland, Oregon. The study aimed to evaluate the level and variations in the quality of preventive and developmental services for young children and assess the contribution of key system, provider and patient factors.

3) HRSA Evaluation Study - The specific goal of this study was to evaluate the feasibility, acceptability and impact of three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young children in the context of discussions between pediatric clinicians and parents during well-child visits. The evaluation measures used data from 5 different tools/surveys including PHDS. The parent-completed PHDS was administered before and after the

intervention to assess changes in the quality of well-child care. The study funded by Health Resources and Services and Administration's (HRSA) Maternal and Child Health Bureau. (Patient Centered Quality Improvement of Well-Child Care, Final Report, Supported by a grant from the Maternal and Child Health Bureau Research Grants Program, Health Resources and Services Administration, R40 MC08959 03-00.)

STUDY POPULATION

Online PHDS: Children age 3-48 months of age whose parents completed the online publicly available PHDS were included in the testing. During 2008-2016, we received 5,670 completed surveys. Of those surveys, 5,355 surveys with provider IDs were used for analyses. Children's socio-demographic and health characteristics varied across the individual providers included in the analysis.

Table 1B.4a: Characteristics of children for whose visited provider ID is available

Characteristics (n=5355)	All Children	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)					
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)	
Age of child							
Under 10 months of age	38.3%	19.1%	49.5%	33.3%	54.4%	24.7%	
10 to 18 months of age	34.7%	39.4%	38.5%	38.9%	29.1%	57.1%	
19-47 months of age	27.0%	41.5%	12.1%	27.8%	16.5%	18.2%	
Race/ethnicity of child							
White, non-Hispanic	53.8%	13.3%	81.0%	20.3%	50.7%	17.3%	
Hispanic	40.8%	81.1%	14.3%	74.7%	40.8%	78.7%	
Other race/ethnicity	5.3%	5.5%	4.7%	7.0%	8.4%	4.0%	
Respondent education level							
Did not complete high school	12.1%	23.6%	0	34.1%	6.4%	15.8%	
Completed high school	88.9%	76.4%	100%	65.9%	93.6%	84.2%	
Children who qualify for Children with Special Health Care Needs (CSHCN) Screener criteria							
CSHCN	10.1%	7.4%	8.8%	10.0%	11.4%	5.2%	
Non-CSHCN	89.9%	92.6%	91.2%	90.0%	88.6%	94.8%	
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	22.7%	-	24.4%	-	28.9%	0%	

-Data is not available due to small sample size

KPNW Study: The population studied was children 3 to 48 months old who live in a metropolitan area in the Pacific Northwest. One randomly selected child per household whose age would be no younger than 3 months of age and no older than 48 months of age at the time that their parents received the survey and had one or more well-child visits were eligible to be sampled. A random sample of 5,755 children were identified. Of the 5,755 sampled children, 2,173 surveys were returned (37.8%).

Table 1B.4b: Characteristics of children for whom survey responses were received, KPNW study, Top 5 individual providers with highest number of surveys

Characteristics (n=2173)	All Children	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)					
		7 (n=80)	53 (n=77)	4 (n=74)	1 (n=67)	43 (n=66)	
Age of child							
Under 10 months of age	22.0%	20.0%	19.5%	20.3%	22.4%	21.2%	
10 to 18 months of age	26.6%	25.0%	29.9%	35.1%	22.4%	15.2%	
19-47 months of age	51.4%	55.0%	50.6%	44.6%	55.2%	63.6%	
Gender of child							
Female child	46.2%	48.8%	49.4%	47.3%	41.8%	45.5%	
Male child	53.8%	51.3%	50.6%	52.7%	58.2%	54.5%	
Race/ethnicity of child							
White, non-Hispanic	72.9%	84.8%	77.0%	93.2%	76.9%	62.5%	
Asian, non-Hispanic	7.8%	2.5%	6.8%	1.4%	3.1%	20.3%	
Hispanic	8.9%	6.3%	12.2%	2.7%	10.8%	10.9%	
Other race/ethnicity	10.4%	6.3%	4.1%	2.7%	9.2%	6.3%	

Child is the first born in the family 52.1% 52.5% 40.8% 35.1% 54.5% 52.3%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS) 31.3% 21.5% 24.7% 27.0%
29.7% 26.2%
Education level of mother
High school or less 12.7% 20.3% 3.9% 14.9% 16.7% 6.2%
More than high school 87.3% 79.7% 96.1% 85.1% 83.3% 93.8%

HRSA Evaluation Study: The study inclusion criteria were used to determine which parents/guardians of children were invited to participate in the interventions and/or evaluation from each participating study site:

- Parent has a well-child visit scheduled at this intervention site for one or more of their children.
- The child is scheduled for their 4-month to 3-year-old well-child visit and, therefore, is between the ages of 4 and 40 months (e.g. 40 month old children could be there for their 3 year well-child visit)
- The parent can read and understand English and is able to complete the intervention and evaluation tools.
- For intervention, the parent was able to access the online version of the Plan My Child's Well-Visit tool and the online evaluation survey.

The analysis includes 551 completed surveys at baseline (2010) and 275 completed surveys at follow-up (2011-12)

Table 1B.4c. Sample description for baseline and follow-up PHDS respondents

	Baseline (n=551)	Follow-up (n=275)
Visit type of child for whom survey was completed		
4, 6 or 9-month	38.9%	36.2%
12, 15 or 18-month	33.7%	41.3%
24 or 36-month	27.4%	22.4%
Birth order of child for whom survey was completed		
First child	42.2%	56.6%
Not first child	57.8%	43.4%
Race/ethnicity		
White, non-Hispanic	80.3%	83.5%
Hispanic	8.4%	6.6%
Other/multiple, non-Hispanic	8.6%	6.6%
Asian, non-Hispanic	2.7%	3.3%
Insurance type		
Private or private and public	90.7%	86.7%
Public only (includes Medicaid, Medicare, CHIP and Military)	7.6%	12.1%
Other	0.7%	0.4%
None	0.9%	0.8%

DISPARITIES

Online PHDS: Variation is observed according to a child's age; race/ethnicity; level of risk for developmental, behavioral, or social delays across all quality measures. For the Anticipatory Guidance and Parental Education measure, care for children 19-48 months, non-Hispanic white children, children whose parents completed at least high school education is most likely to meet scoring criteria.

Table 1B.4D. Anticipatory guidance and parental education by child demographics and other characteristics

Characteristics	All children n	%		
Age groups				
3-8 months	1347	60.4%		
9-18 months	1104	57.4%		
19-48 months	889	63.1%		
p values (Pearson chi-square)		-	0.003	
Gender				
Male	372	59.7%		
Female	364	57.7%		
p values (Pearson chi-square)		-	0.48	
Race/ethnicity				

Hispanic	1150	58.9%						
White non-Hispanic	1735	61.9%						
Black non-Hispanic	58	56.9%						
Asian non-Hispanic	47	42.7%						
Other/Multi race, non-Hispanic	52	60.5%						
p values (Pearson chi-square)	-	<0.0001						
Adult survey responds education level								
Did not complete high school	341	55.3%						
Completed high school or higher education	2879	60.9%						
p values (Pearson chi-square)	-	0.01						
CSHCN status								
Non-CSHCN	3002	60.1%						
CSHCN	338	59.3%						
p values (Pearson chi-square)	-	0.70						
At risk for developmental delay (online only)								
Low/No risk	1394	63.0%						
High/Moderate risk	423	56.2%						
p values (Pearson chi-square)	-	0.001						

KPNW study: After controlling for other child and family demographic and health factors and provider characteristics, the likelihood (or adjusted odds ratio-AOR) that a child met quality measure criteria differed significantly according to: (1) child's age, (2) child's race/ethnicity, (3) child's birth, (4) child's developmental and behavioral risk status, (5) parent risk for depression for three measures.

Table 1B.4E: Mean number of developmental services care components for which quality care was received and the proportion of children meeting criteria for receiving quality developmental services by characteristics of children and families.

Characteristic of Child or Child's Family	% Meeting all criteria	% Meeting behavioral or developmental topical area	% Meeting physical care topical area criteria	% Injury prevention topical area criteria
Child's Age				
Less than 9 mos.	37.5%ns	49.6% S	47.9% S	70.0% S
10 to 18 mos.	37.1%	48.0%	51.8%	55.9%
		AOR: 1.39		AOR: .59
19 to 49 mos.	42.0%	56.5%	62.0%	55.4%
	AOR: 1.42		AOR: 2.01	AOR: .50
Child's Gender				
Male Child	39.5% NS		52.2% NS	56.2% NS 57.9% NS
Female Child	39.8%	53.2%	56.2%	59.8%
Child's Race				
White, Non-Hispanic		41.8%s	55.8% S	59.4% S 61.5% S
Asian, Non-Hispanic		31.5%	39.3%	42.00% 51.20%
	AOR: .54	AOR: .57		
Hispanic 33.2%	46.5%	51.6%	51.10%	
37.2%	AOR: .67			AOR: .60
Other Race,		47.9%	50.9%	53.5%
Multiple Race				AOR: .72
Birth Order				
Not First Born	51.5% S	65.8% S	67.7% S	68.4% S
First Born	28.9%	40.8%	45.8%	50.0%
	AOR: .33	AOR: .39	AOR: .43	
Child's Risk for Developmental, Behavioral or Social Delays (Using Parent's Evaluation of Developmental Status)				
Low/No Risk	45.6%s	59.3% S	62.9% S	64.2% S
At Risk 27.1%	38.6%	41.6%	46.9%	
	AOR: .45	AOR: .42	AOR: .57	
Respondent Education				
More than High School	39.0%s	52.4% NS	56.0% NS	57.9% S

High School or Less	45.7%	56.1%	60.5%	66.4%
Respondent's Risk for Depression (Using the Kemper Screener)				
No Symptoms of Depression	42.1% S	55.5% S	58.7% S	60.8% S
Symptoms of Depression	32.4%	38.2%	45.1%	46.1%
	AOR: .49	AOR: .59	AOR: .54	

NOTE: Adjusted odds ratios (AOR) derived from regression analyses listed in the table are shown only if they are statistically significant. AOR uses the first subgroup of each characteristic as a reference. s=differences significant at the p < .05 level of significance; NS=differences not significant.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4**

NA

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

De.7. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** CAHMI_Data_Dictionary_Anticipatory_Guidance_PHDS_01_11_17.pdf

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.
No

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.
NA

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the*

rationale for the measure.
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator is the number of parents who had a well child visit within the last 12 months and who indicated that they received anticipatory guidance and education, that their questions were answered or that they already had the information and did not require anticipatory guidance on that topic.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator is the number of parents who responded to the AGPE items as either (1) "Yes, and my questions were answered" OR "(2) Yes, but my questions were not answered completely" responses to items in the "Anticipatory Guidance and Parental Education (AGPE)" scale.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Parents whose children ages 0-48 months who received a well-child visit in the last 12 months and who responded to at least half of the AGPE items (see Attachment A-2 pages 8-10) on the Promoting Healthy Development Survey (PHDS: www.wellvisitsurvey.org)

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Parents whose children ages 0-48 months who received a well-child visit in the last 12 months and who responded to at least half of the AGPE items (see Attachment A-2 pages 8-10) on the Promoting Healthy Development Survey (PHDS: www.wellvisitsurvey.org)

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Unknown and missing values (responses coded missing) are excluded in the data analysis. Approximately 2.6% of parents who started the Online PHDS did not complete the survey (range 0.0-3.3% for top 5 providers with highest number of surveys; see Testing form, pages 23-24 for more detailed information on missing data).

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
Unknown and missing values (responses coded missing) are excluded in the denominator for the data analysis. If a parent answered less than half of the items, their data are considered to be missing and were excluded from analysis. THE AGPE measure had less than 2% of missing cases, ranging from 0-3.3% caross the top 5 providers with the highest number of surveys. Few overall missing values sugest that the measure level results are unlikely to be biased by non-response to the survey questions

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*
Although no stratification is required, the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the findings by possible vulnerability, should any individual provide have sufficient data (parent responses) to do so. Potential variables for stratification include:
(1) Child demographic characteristics (e.g., the child´s age, race);
(2) Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or
(3) Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
No risk adjustment or risk stratification
If other:

S.12. **Type of score:**
Rate/proportion
If other:

S.13. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
The numerator is the sum of survey respondents (parents) answering either "Yes and my questions were answered" or "No, but I already had information about this topic and did not need to talk about it anymore." A score of at least 75% aggregated across all AGPE items (n=15 or 16 depending on child age) represents quality for anticipatory guidance. See also Attachment A-4.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
Not applicable

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Data are collected using the parent-reported "Promoting Healthy Development Survey" (PHDS) developed by the CAHMI (www.wellvisitsurvey.org). Instructions for survey completion are included with the survey. AGPE items are multiple choice (see Attachment A-2 pages 8-10). There are 15 or 16 items depending on the child´s age. The items can be grouped by physical care, development and behavior and injury prevention. The PHDS is initiated by the provider who can send it to all parents who have received a well child visit. CAHMI has a website (www.phdstoolkit.org) where providers can register to use the PHDS. This site assigns each provider a unique URL, which allows for provider identification by CAHMI as well as light branding with the provider´s logo so that it is identifiable by the parent. The PHDS Toolkit website sends an email to the provider with the unique URL link to the survey. The provider then sends the link to the parents asking them to fill out the survey and provide feedback about the visit. The parent fills out the survey and receives a customized feedback report. The survey data are captured on a secure HIPAA compliant CAHMI server. Through the PHDS Toolkit website, providers can generate a report that aggregate parent data from the survey. Providers must have a minimum of 10 surveys to generate a report to maintain parent confidentiality. See Evidence Form, Figure 1 for the logic model.

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Other

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The AGPE measure is included as part the CAHMI Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org, Attachment A-2). The data are generated by parents filling out the PHDS. The PHDS is based in English. See Evidence Form, Figure 1 for a visual model of the data collection process.

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Individual

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Clinician Office/Clinic
If other:

S.22. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
NA

2. Validity – See attached Measure Testing Submission Form
CAHMI_NQF_Testing_Attachment_Anticipatory_Guidance_020217_Final.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)
Yes

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)
No

2.3 For maintenance of endorsement
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care)

What were the statistical results of the analyses used to select risk factors?

Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)
No - This measure is not risk-adjusted

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2958
Measure Title: Anticipatory Guidance and Parental Education
Date of Submission: 2/2/2017
Type of Measure:

☒ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process	☐ Efficiency
☐ Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes
10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).
11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.
12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.
13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.
14. Risk factors that influence outcomes should not be specified as exclusions
15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE
Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry

☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Patient reported data	☒ other: Patient reported data

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

We used the following data sources for testing of the quality measure:

1) Online Promoting Healthy Development Survey (PHDS) – data collected through an online, publicly available tool (Promoting Healthy Development Survey-PHDS). Parents who had a well-child care visit in the last 12 months can complete the PHDS. Providers initiate the survey. (See Evidence Form Figure 1 for the Online PHDS logic model.)
2) Kaiser Permanente Northwest (KPNW) Study – CAHMI partnered with Kaiser Permanente Northwest in Portland, Oregon. The study aimed to evaluate the level and variations in the quality of preventive and developmental services for young children and assess the contribution of key system, provider and patient factors.
3) HRSA Evaluation Study - The specific goal of this study was to evaluate the feasibility, acceptability and impact of three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young children in the context of discussions between pediatric clinicians and parents during well-child visits. The evaluation measures used data from 5 different tools/surveys including PHDS. The parent-completed PHDS was administered before and after the intervention to assess changes in the quality of well-child care. The study funded by Health Resources and Services and Administration's (HRSA) Maternal and Child Health Bureau. (Patient Centered Quality Improvement of Well-Child Care, Final Report, Supported by a grant from the Maternal and Child Health Bureau Research Grants Program, Health Resources and Services Administration, R40 MC08959 03-00.)

1.3. What are the dates of the data used in testing? 2004-2016
Online PHDS: 2008-2016
KPNW Study: 2004-2005
HRSA Evaluation Study: 2010-2012

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Online PHDS: n=5,670 surveys reporting on quality of care provided by 299 individual pediatricians and primary care providers from 88 clinics in 36 states. Participation is a voluntary self-selection process based on knowledge and interest in quality improvement in their practice.

KPNW Study: Provider-level surveys and quality of care assessment were focused on the care provided by 56 individual providers (44 pediatricians, 9 nurse practitioners, 3 physician assistants) in the pediatrics department who were organized into ten geographically distinct offices.

HRSA Evaluation Study: Three pediatric offices in Oregon: 1) a rural site, (4 pediatricians), 2) an urban site (8 pediatricians), and 3) an urban site, (12 pediatricians). All pediatricians in selected clinic and office staff participated in relevant baseline and follow up data collection.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

<u>Online PHDS:</u> Children age 3-48 months of age whose parents completed the online publicly available PHDS were included in the testing. During 2008-2016, we received 5,670 completed surveys. Of those surveys, 5,355 surveys with provider IDs were used for analyses. Children's socio-demographic and health characteristics varied across the individual providers included in the analysis.

Table 1.6a: Characteristics of children for whose visited provider ID is available

Characteristics	All Children (n=5355)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
Age of child						
Under 10 months of age	38.3%	19.1%	49.5%	33.3%	54.4%	24.7%
10 to 18 months of age	34.7%	39.4%	38.5%	38.9%	29.1%	57.1%
19-47 months of age	27.0%	41.5%	12.1%	27.8%	16.5%	18.2%
Race/ethnicity of child						
White, non-Hispanic	53.8%	13.3%	81.0%	20.3%	50.7%	17.3%
Hispanic	40.8%	81.1%	14.3%	74.7%	40.8%	78.7%
Other race/ethnicity	5.3%	5.5%	4.7%	7.0%	8.4%	4.0%
Respondent education level						
Did not complete high school	12.1%	23.6%	0	34.1%	6.4%	15.8%
Completed high school	88.9%	76.4%	100%	65.9%	93.6%	84.2%
Children who qualify for Children with Special Health Care Needs (CSHCN) Screener criteria						
CSHCN	10.1%	7.4%	8.8%	10.0%	11.4%	5.2%
Non-CSHCN	89.9%	92.6%	91.2%	90.0%	88.6%	94.8%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	22.7%	-	24.4%	-	28.9%	0%

-Data are not available due to small sample size.

KPNW Study: The population studied was children 3 to 48 months old who live in a metropolitan area in the Pacific Northwest. One randomly selected child per household whose age would be no younger than 3 months of age and no older than 48 months of age at the time that their parents received the survey and had one or more well-child visits were eligible to be sampled. A random sample of 5,755 children were identified. Of the 5,755 sampled children, 2,173 surveys were returned (37.8%).

Table 1.6b: Characteristics of children for whom survey responses were received, KPNW study, Top 5 individual providers with highest number of surveys

Characteristics	All Children (n=2173)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		7 (n=80)	53 (n=77)	4 (n=74)	1 (n=67)	43 (n=66)
Age of child						
Under 10 months of age	22.0%	20.0%	19.5%	20.3%	22.4%	21.2%
10 to 18 months of age	26.6%	25.0%	29.9%	35.1%	22.4%	15.2%
19-47 months of age	51.4%	55.0%	50.6%	44.6%	55.2%	63.6%
Gender of child						
Female child	46.2%	48.8%	49.4%	47.3%	41.8%	45.5%
Male child	53.8%	51.3%	50.6%	52.7%	58.2%	54.5%
Race/ethnicity of child						
White, non-Hispanic	72.9%	84.8%	77.0%	93.2%	76.9%	62.5%
Asian, non-Hispanic	7.8%	2.5%	6.8%	1.4%	3.1%	20.3%
Hispanic	8.9%	6.3%	12.2%	2.7%	10.8%	10.9%
Other race/ethnicity	10.4%	6.3%	4.1%	2.7%	9.2%	6.3%
Child is the first born in the family	52.1%	52.5%	40.8%	35.1%	54.5%	52.3%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	31.3%	21.5%	24.7%	27.0%	29.7%	26.2%
Education level of mother						
High school or less	12.7%	20.3%	3.9%	14.9%	16.7%	6.2%
More than high school	87.3%	79.7%	96.1%	85.1%	83.3%	93.8%

HRSA Evaluation Study: The study inclusion criteria were used to determine which parents/guardians of children were invited to participate in the interventions and/or evaluation from each participating study site:

- Parent has a well-child visit scheduled at this intervention site for one or more of their children.
- The child is scheduled for their 4-month to 3-year-old well-child visit and, therefore, is between the ages of 4 and 40 months (e.g. 40 month old children could be there for their 3 year well-child visit)
- The parent can read and understand English and is able to complete the intervention and evaluation tools.
- For intervention, the parent was able to access the online version of the Plan My Child's Well-Visit tool and the online evaluation survey.

The analysis includes 551 completed surveys at baseline (2010) and 275 completed surveys at follow-up (2011-12)

Table 1.6c. Sample description for baseline and follow-up PHDS respondents

	Baseline (n=551)	Follow-up (n=275)
Visit type of child for whom survey was completed		
4, 6 or 9-month	38.9%	36.2%
12, 15 or 18-month	33.7%	41.3%
24 or 36-month	27.4%	22.4%
Birth order of child for whom survey was completed		
First child	42.2%	56.6%
Not first child	57.8%	43.4%
Race/ethnicity		
White, non-Hispanic	80.3%	83.5%
Hispanic	8.4%	6.6%
Other/multiple, non-Hispanic	8.6%	6.6%
Asian, non-Hispanic	2.7%	3.3%
Insurance type		
Private or private and public	90.7%	86.7%
Public only (includes Medicaid, Medicare, CHIP and Military)	7.6%	12.1%
Other	0.7%	0.4%
None	0.9%	0.8%

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Online PHDS and KPNW study data were used for reliability testing and stratification analysis. Validity findings are presented from a peer-reviewed publications and online PHDS. Performance analysis was conducted using the online PHDS, KPNW study and HRSA Evaluation Study data.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Online PHDS: Child's age, sex, race/ethnicity, and respondent (parent) age, race/ethnicity, and education level. The survey does not have a question asks about family income due to complexity of collecting income data by self-reported survey. However, the online PHDS has items assessing the family's economic situation: How much trouble does the family have paying for a) child's health and medical expenses; b) supplies like formula, food, diapers, clothes and shoes; and c) health care for the parent.
KPNW Study: Child's age, sex, race/ethnicity, and education level of mother
HRSA Study: Child's age, race-ethnicity, and insurance type

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Internal consistency: Cronbach's alpha was calculated to assess internal consistency of survey items used in the Anticipatory Guidance and Parental Education (APGE) measure. Cronbach's alpha is the most widely used in health care research when multiple-item measures of a concept or construct are employed. The acceptable values of alpha ranges from 0.70 to 0.95. Survey questions on AGPE are age specific. Therefore, testing was done by children's age group (3-9, 10-18 and 19-48 months). The APE questions can be also grouped by topical area: Physical Care, Developmental and Behavioral Guidance and Injury Prevention.

The primary aim of the AGPE quality measure is to detect the difference between providers on the quality of care provided to young children. Variance between and within group (provider) were calculated using ANOVA. Then we calculated intra-class correlation (ICC) as a ratio of the variance between groups over the total variance. The interpretation of the ICC is as the proportion of relevant variance that is associated with differences among measured objects.[1] Fleiss (1981) and Cicchetti and Sparrow (1981) from the medical group state that ICC range categories are: < 0.40 = poor; 0.40 – 0.59 = fair; 0.60 – 0.74 = good; and > 0.74 = Excellent[2]. Values above about 0.7-0.8 are considered acceptable for applied tests. In the analysis, we included providers with 20 or more surveys.

1. McGraw, K. O. and Wong, S. P. Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1996:1(1), 30-46.
2. Cicchetti D.V. and Sparrow, S.S. Developing criteria for establishing the interrater reliability of specific items in a given inventory. American Journal of Mental Deficiency, 1981:86, 127-137.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 2a2.3a. Anticipatory Guidance and Parental Education: Content, Scoring and Internal Consistency, Online PHDS, all providers and top 5 individual providers with highest number of surveys

What is measured	Scoring	Internal Consistency (Cronbach's Alpha)					
Four multi-part items assess whether general and age specific anticipatory guidance topics are addressed. Includes feeding and nutrition, sleeping and physically caring for child, safety and injury prevention, child growth, development, communication and behavior	Mean score on a multi-item scale	All providers (n=5355)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
			1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
By age group:							
- 3-9 months		0.90*	0.98*	0.89*	0.89*	0.76*	0.95*
- 10-18 months		0.89*	0.87*	0.86*	0.94*	0.85*	0.85*
- 19-48 months		0.90*	0.80*	0.71*	0.81*	0.98*	0.88*
By topical area:							
- Physical		0.77*	-	-	-	-	-
- Behavioral or developmental		0.86*	-	-	-	-	-
- Injury prevention		0.73*	-	-	-	-	-

*Met criteria for reliability and internal consistency.

- Data not available due to small sample size

Table 2b. Anticipatory Guidance and Parental Education: Content, Scoring and Internal Consistency, KPWN study, all providers and top 5 individual providers with highest number of surveys

What is measured	Scoring	Internal Consistency (Cronbach's Alpha)					
Four multi-part items assess whether general and age specific anticipatory guidance topics are addressed. Includes feeding and nutrition, sleeping and physically caring for child, safety and injury prevention, child growth, development, communication and behavior	Mean score on a multi-item scale	All providers	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
			7 (n=80)	53 (n=77)	4 (n=74)	1 (n=67)	43 (n=66)
By age group:							
- 3-9 months		0.90*	0.87*	0.83*	0.93*	0.97*	0.93*
- 10-18 months		0.91*	0.89*	0.92*	0.89*	0.85*	0.93*
- 19-48 months		0.89*	0.88*	0.87*	0.83*	0.87*	0.85*
By topical area:							
- Physical		0.71*-0.80*	-	-	-	-	-
- Behavioral or developmental		0.80*-0.85*	-	-	-	-	-
- Injury prevention		0.65-0.71*	-	-	-	-	-

*Met criteria for reliability and internal consistency.

Chronbach's alpha ranged 0.71-0.98 across providers, age groups and topical areas of the anticipatory guidance items (Online PHDS). These findings are consistent with the findings of the previous peer-reviewed publications.[3,4] Intraclass correlation coefficient for the Anticipatory Guidance and Parental Education measure is 0.72, indicating that 71,5% of the variance in the mean of the providers is "true" rather than due to chance.

3. Bethell C, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. Pediatrics. 2001 May;107(5):1084-94.
4. Christina Bethell, PhD, MPH, MBA; Colleen H. Peck Reuland, MS; Neal Halfon, MD, MPH; Edward L. Schor, Measuring the Quality of Preventive and Developmental Services for Young Children: National Estimates and Patterns of Clinicians' Performance. Pediatrics, 2004, 113(6):1973-83

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Psychometric item level testing demonstrated that the Anticipatory Guidance and Parental Education quality measure provides psychometrically reliable assessment of the provision of nationally recommended well-child care with strong internal consistency (Cronbach's alpha ranges 0.71-0.98 across age-specific and topical areas, individual providers) and good intraclass correlation (0.72).

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

A standard, multistage process was used to ensure validity of the AGPE measure:

- Focus groups and in-depth cognitive interviews were conducted throughout the survey development process;
- A review of literature identified through Medline or during key informant interviews; and,
- Three Advisory Groups comprised of pediatricians, family practitioners, consumer representatives, public health experts, and researchers, regularly reviewed and provided input on the identification of quality measurement topics and the development of the PHDS.

A "gold standard" does not exist for determining the criterion validity of patient-reported measures of quality. However, to ensure the validity of the AGPE quality measure results, we followed rigorous procedures representing best practices within the field to develop the survey questions. To ensure the content validity of measures of parent experiences, we used qualitative methods, including both focus groups and cognitive interviews, to inform development and evaluation of the AGPE questions.
Focus groups with families aimed to identify the aspects of health care quality that are important to parents in the area of preventive care for their children. In-depth cognitive testing of the draft survey items was conducted with 15 families representing a range of racial, income and education groups as well as different types of health insurance coverage, age of child, age and sex of parent, and number of children in family. Focus groups and cognitive interviews with 35 health care providers in Vermont and Washington and 20 parents of young children in Vermont were conducted to inform item-reduction, administration specifications, and reporting templates. Survey modifications were made based on findings in order to improve the reliability, validity and cognitive ease of the AGPE items.

Factor analysis was conducted to assess the construct validity of the AGPE quality measure. Each of the survey items used to construct the PHDS scale-based quality measures were used in the factor analysis.[1] Acceptable level of factor loading for instruments developed for research purposes can be as low as 0.60[2] and factor loading more than this threshhold is considered as a strong association.[3] Pearson correlation coefficients were calculated between age-specific anticipatory guidance scales to assess the degree to which each of the item provide unique information.

To assess the concurrent validity of the measure scale, hypothesized associations among PHDS items were examined using logistic regression model. Three hypotheses were evaluated:

Respondents who indicate that providers talked with them about recommended anticipatory guidance topics are:

- more likely to report increased confidence as a parent because of interactions with health care providers
- more likely to report positive parenting behaviors in related areas
    - less likely to report being concerned about their child's development in related areas compared with respondents who indicate that providers did not talk with them although they wished they had done so.[1]

Pearson correlation coefficients were calculated across all scale measures to test hypotheses about expected relationships among the PHDS quality measures and to assess the degree to which each of the PHDS quality measures provide unique information. We expect a moderate or strong correlation between the family assessment scale measures (>0.30) and inter-scale correlation coefficients to be less than 0.80.

1. Bethell C, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. Pediatrics. 2001 May;107(5):1084-94.
2. Suhr D and Shay M. Guidelines for reliability, confirmatory and exploratory factor analysis. Accessed at: http://www.wuss.org/proceedings09/09WUSSProceedings/papers/anl/ANL-SuhrShay.pdf. Retrieved 02/01/2017
3. Costello A.B and Osborne J.W. Best Practices in Exploratory Factor Analysis: Four recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation. 2005:10(7). Accessed at: http://www.pareonline.net/pdf/v10n7.pdf, Retrieved 02/01/2017

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Using behavior coding methods, for each item in the AGPE quality measure, instances where the respondent required clarification or did not appropriately answer an item were noted. Also, items where the interviewer had difficulty asking the question without edits to the wording were noted. Data analysis was used to inform item-reduction. Content was revised and refined iteratively with each set of interviews.

Cognitive testing confirmed the readability of the AGPE items for people across a range of educational levels. Parents were uniformly able to complete the PHDS self-administered survey in 10-15 minutes. Readability assessments indicated the AGPE items to be written at the 8th-9th grade reading level. Survey design and formatting were finalized with input from a group of experts and family representatives.

Factor analysis demonstrated a strong factor structure within the PHDS measures. Average factor loading for anticipatory guidance was 0.72. Factor analysis suggests that the scale items are unidimensional. The strength of the observed inter item correlations were not so high as to suggest redundancy across items (average correlation: 0.36). The highest correlation observed was between the "what your child is able to understand " and the "how your child communicates his/her needs" items (0.67).

Concurrent validity testing showed that parents reporting positive parenting behaviors had significantly higher scores on the anticipatory guidance quality measure compared with parents not reporting positive behaviors. Parents who reported that their questions on specific anticipatory guidance topics were answered were more likely to report higher confidence in related parenting activities because of information and counseling received from their child's doctor or other health care providers compared with parents answering "no, but I wish we had discussed that" (odds ratio [OR]: 5.9, 95% confidence interval [CI]: 3.4-10.2; OR: 8.3, 95% CI: 5-13.8). Moreover, parents who reported positive parenting behaviors in the areas of injury prevention (70.9 vs 92.1, $P < .000$) and reading to their child (69.1 vs 13, $P < .000$) also had significantly higher scores on the "anticipatory guidance from providers" quality measure. In addition, significantly fewer parents reported concerns about their child's behavior if they also reported that their child's doctor or other health care providers talked with them about the kinds of behaviors they might expect to see in their child (46.7% "yes, talked" vs 65.5% "no, wish", $P < .000$; OR: 0.46 95% CI: 0.29 – 0.72).

Correlations between the PHDS quality measures were not so high as to suggest redundancy across measures (average correlation: 0.34). The highest correlation observed was between the "Assessment of family psychosocial well-being" & "Assessment of smoking, drug and alcohol use and safety in the family" (0.54) and "anticipatory guidance from providers" & the "family-centered care" measures (0.52) as expected.

Table 2b2.3a. Pearson Correlation Coefficients among PHDS Quality Measures (online PHDS)

Scale Measures	Anticipatory Guidance and Parent Education	Family Centered Care	Ask About Parental Concern	Assessment of smoking, drug and alcohol use and safety in the family	Assessment of family psychosocial well-being
Family Centered Care	.52				
Ask About Parental Concern	.16	.14			
Assessment of smoking, drug and alcohol use and safety in the family	.16	.13	.07		
Assessment of family psychosocial well-being	.19	.16	.09	.54	

Average correlation: 0.34

Most of the AGPE items have been used in the National Survey on Early Childhood Health. The AGPE quality measures are among the few recognized in the Agency for Healthcare Research and Quality's Child Health Toolbox and the National Quality Measures Clearinghouse as measures that meet basic criteria for use as standardized indicators of health care quality for children.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The AGPE quality measure provides conceptually and psychometrically valid assessment of the provision of nationally recommended preventive care services for young children, with strong construct validity (average factor loading: 0.72). Each of the PHDS quality measure provides unique information about performance. The measure is used in national surveys and recognized as measures that meet basic criteria for use as standardized indicators of health care quality for children. The measure serves as an important complement to existing quality measures.

_____

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Not applicable

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Not applicable

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is*

transparent, e.g., scores with and without exclusion)

Not applicable

_____

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification** **NOTE No Risk Adjustment only**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☒ **Stratification by** variable number of **risk categories**

☐ **Other,**

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

Not applicable

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

The AGPE quality measure does not require risk adjustment because we do not expect variation in the quality of care provided for children due to risk factors, e.g. children with special health care needs. The performance should be the same regardless of risk factors. The national experts extensively reviewed the risk adjustment requirements during development of the AGPE items and composite measure and did not recommend risk-adjustment for the measures. In addition, during the KPNW study, we assessed whether the probability of receiving guidance, education or screening was higher according to a child's level of need or risk, thereby indicating that providers are customizing care to children. The study found no evidence that providers customize care to children most at risk.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Identification of variation in AGPE quality measure across subgroups of children helps to highlight aspects of care and population of children for which preventive and developmental services may be most need of improvement. Although no stratification is required (number of surveys for each individual providers may not be sufficient to stratify), the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the quality measures by possible vulnerability:

- Child demographic characteristics (e.g., the child's age, race)
- Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs)
- Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

Based on extensive literature review and expert panel, we identified that child and parent demographics such as age, sex, race-ethnicity, income, insurance, parent behavior, CSHCN screener and follow-up for children at risk can be used for stratification. Several studies have documented differences in access and quality of care provided to children, as well as in parent-reported satisfaction with care.[1-2] One study found: "Overall, 94.0% of parents reported 1 or more unmet needs for parenting guidance, education, and screening by pediatric clinician(s) in 1 or more of the content of care areas evaluated. Uninsured children and children aged 18 to 35 months are disproportionately represented among the 15.3% of children whose parents indicated an unmet need in each of the 4 areas of care. There are significant variations in

performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors. "[3]

The KPNW study assessed child and family characteristics to characterize the child and their family based on the PHDS item responses: child's race/ethnicity, birth order, risk for developmental, behavioral, or social delays using responses to Frances Glascoe's Parents' Evaluation of Developmental Status (PEDS) items included in the ProPHDS 29 parent's education; and whether he/she is experiencing symptoms of depression using Kathy Kemper's screening items. Adjusted odds ratios were calculated using logistic regression analysis in order to assess differences in the odds of meeting quality measure criteria according to child, family and provider characteristics, after controlling for other variables.

References:
1. Halfon N, Regalado M, Sareen H,Inkelas M, Reuland CH, Glascoe FP,Olson LM. Assessing development in the pediatric office. Pediatrics. 2004 Jun;113(6 Suppl):1926-33.
2. Weech-Maldonado R, Morales LS, Spritzer K, Elliott M, Hays RD. Racial and ethnic differences in parents' assessments of pediatric care in Medicaid managed care. Health Serv Res. 2001 Jul;36(3):575-94.
3. Bethell C, Reuland CH, Halfon N, Schor EL. Measuring the quality of preventive and developmental services for young children: national estimates and patterns of clinicians' performance. Pediatrics. 2004 Jun;113(6 Suppl):1973-83.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

See 2b4.3.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Pearson's chi-squire test was used to compare the prevalence of AGPE quality measure across the stratification characteristics. We preformed logistic regression analysis in order to assess differences in the odds of meeting quality measure criteria according to child, family and provider characteristics, after controlling for other variables.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable

2b4.9. Results of Risk Stratification Analysis:

Online PHDS: Variation is observed according to a child's age; race/ethnicity; level of risk for developmental, behavioral, or social delays. For the Anticipatory Guidance and Parental Education measure, care for children 3-48 months, non-Hispanic white children, children whose parents completed at least high school education is most likely to meet scoring criteria. Care for children with low/no risk for developmental delay more likely to meet the AGPE quality measure compared to children with high/medium risk.

Table 2b4.9a. Anticipatory guidance and parental education by child demographics and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	1347	60.4%
9-18 months	1104	57.4%
19-48 months	889	63.1%
p values (Pearson chi-square)	-	0.003
Gender		
Male	372	59.7%
Female	364	57.7%
p values (Pearson chi-square)	-	0.48
Race/ethnicity		
Hispanic	1150	58.9%
White non-Hispanic	1735	61.9%
Black non-Hispanic	58	56.9%
Asian non-Hispanic	47	42.7%
Other/Multi race, non-Hispanic	52	60.5%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	341	55.3%
Completed high school or higher education	2879	60.9%
p values (Pearson chi-square)	-	0.01
CSHCN status		
Non-CSHCN	3002	60.1%
CSHCN	338	59.3%
p values (Pearson chi-square)	-	0.70
At risk for developmental delay (online only)		
Low/No risk	1394	63.0%
High/Moderate risk	423	56.2%
p values (Pearson chi-square)	-	0.001

KPNW study: After controlling for other child and family demographic and health factors and provider characteristics, the likelihood (or adjusted odds ratio-AOR) that a child met quality measure criteria differed significantly according to: (1) child's age, (2) child's race/ethnicity, (3) child's birth order, (4) responded education level, (5) child's developmental and behavioral risk status, and (6) parent risk for depression.

Table 2b4.9b: Mean number of developmental services care components for which quality care was received and the proportion of children meeting criteria for receiving quality developmental services by characteristics of children and families.

Characteristic of Child or Child's Family	% Meeting all criteria	% Meeting behavioral or developmental topical area	% Meeting physical care topical area criteria	% Injury prevention topical area criteria
Child's Age				
Less than 9 mos.	37.5%[ns]	49.6% [S]	47.9% [S]	70.0% [S]
10 to 18 mos.	37.1%	48.0%	51.8% AOR: 1.39	55.9% AOR: .59
19 to 49 mos.	42.0%	56.5% AOR: 1.42	62.0% AOR: 2.01	55.4% AOR: .50
Child's Gender				
Male Child	39.5% [NS]	52.2% [NS]	56.2% [NS]	57.9% [NS]
Female Child	39.8%	53.2%	56.2%	59.8%
Child's Race				
White, Non-Hispanic	41.8%[s]	55.8% [S]	59.4% [S]	61.5% [S]
Asian, Non-Hispanic	31.5%	39.3% AOR: .54	42.00% AOR: .57	51.20%
Hispanic	33.2% 37.2%	46.5% AOR: .67	51.6%	51.10% AOR: .60
Other Race, Multiple Race		47.9%	50.9%	53.5% AOR: .72
Birth Order				
Not First Born	51.5% [S]	65.8% [S]	67.7% [S]	68.4% [S]
First Born	28.9%	40.8% AOR: .33	45.8% AOR: .39	50.0% AOR: .43
Child's Risk for Developmental, Behavioral or Social Delays (Using Parent's Evaluation of Developmental Status)				
Low/No Risk	45.6%[s]	59.3% [S]	62.9% [S]	64.2% [S]
At Risk	27.1%	38.6% AOR: .45	41.6% AOR: .42	46.9% AOR: .57
Respondent Education				
More than High School	39.0%[s]	52.4% [NS]	56.0% [NS]	57.9% [S]
High School or Less	45.7%	56.1%	60.5%	66.4%
Respondent's Risk for Depression (Using the Kemper Screener)				
No Symptoms of Depression	42.1% [S]	55.5% [S]	58.7% [S]	60.8% [S]
Symptoms of Depression	32.4%	38.2% AOR: .49	45.1% AOR: .59	46.1% AOR: .54

NOTE: Adjusted odds ratios (AOR) derived from regression analyses listed in the table are shown only if they are statistically significant. AOR uses the first subgroup of each characteristic as a reference.
[s] differences significant at the $p < .05$ level of significance.
[NS] differences not significant.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

The demographic and socioeconomic survey items included in the AGPE quality measure make it possible for providers to identify populations and subgroups for which health service delivery improvement is most needed.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)
Not applicable

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Differences in the AGPE quality measure across providers is demonstrated for (1) 5 top individual providers with the highest number of surveys using Online PHDS data; (2) across 56 providers using KPNW study data; and (3) pre-post changes across time (2010-2012) after small invervention using HRSA study data for an illustrative purpose.

Online PHDS: The performance scale for the AGPE quality measure was calculated using the scoring methods described in Attachment A-4. Individual provider level differences in performance were illustrated by the proportion of children meeting the quality of care criteria across 5 top providers with the highest number of completed surveys after their well-child visit.

KPNW Study: The significance of differences observed in the proportion of children meeting criteria for the AGPE quality measure across pediatric providers (n=56) was evaluated using t-tests. The relative spread in the quality measure score across providers was assessed using the coefficient of variation statistics (standard deviation across providers multiplied by 100%). Multi-level regression models were conducted using the pediatric provider as the level 2 clustering variable, in order to assess the degree to which the probability that a child meets criteria on each quality measure is explained by differences between providers (called the "clustering effect"). In implementing this multi-level regression method (Empty Model), the presence of a significant clustering effect by pediatric providers was estimated prior to accounting for the child and family characteristics associated with each provider. Second, variables related to the child and family characteristics (child's age, gender, race/ethnicity, birth order, developmental and behavioral delay risk status; parent education and risk for depression) were added to the Empty Model to assess how much of the provider clustering effect observed remains after accounting for these characteristics (called the "Patient Model").

HRSA study: Quantitative data results for the baseline (2010) and follow-up (2011-12) study of the intervention sites using the HRSA Evaluation Study data were conducted using basic descriptive statistics to describe each sample and applying chi-square test of statistical significance to assess differences in the quality measure for the baseline and follow-up samples.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Online PHDS: Table 2b5.2a presents the proportion of children whose parents received anticipatory guidance and parental education on at least half of the topics and met criteria for the quality measure across 5 providers. The proportion of parents who reported discussion of all anticipatory guidance and parental education topics or reported no need of discussion among unaddressed topics ranged 46.8-84.8% across 5 observed providers.

Table 2b5.2a: Proportion of Children Meeting Measure Criteria, Top 5 individual providers with highest number of surveys

Characteristics	All Children (n=5355)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
Children whose parents had their needs met on all items	60.0%	69.6%	47.3%	66.7%	84.8%	46.8%

KPNW Study: The proportion of children whose parents had their anticipatory guidance and parental education informational needs met (meaning they reported either "yes, topic was discussed" or "no, but I already had information and did not need to talk about it" on each topic) was 58.8% on injury prevention topics (17.6% reported "yes, topic was discussed" to all topics), 56.2% on physical care topics (10.4% reported "yes" to all topics) and 52.7% on child development and behavior topics (13.1% reported "yes" to all topics). About two in five of children (39.7%) had parents who reported having their needs met across all topics. Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant (p=0.003).

Table 5b: Proportion of all children in the study who met criteria for receiving quality services and ranges in proportion across providers.

Developmental Services Quality Measures	Proportion of All Children Meeting Measure Criteria (n = 2173)	Range in the Proportion of Children Meeting Measure Criteria Across 51 Pediatric Providers	Relative Variation (COV) in Measure Scores Across Pediatric Providers
Children whose parents had their needs met on all items	39.7%	22.2% to 66.7% SD:10 (p = 0.03)	25.1%

Only providers with n=15 or more responses are included in the provider level analysis. Provider level sample sizes range from 15 to 153.

Multi-level analysis: For the Empty Model that used the provider as the level 2 clustering variable, only 1.1% to 2.2% of the total variance observed in whether children met criteria for all quality measures was explained by either measured or unmeasured differences between the providers that they see. This suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers. These findings translate into a 1.19 to 1.29 median odds ratio across the all quality measures in the PHDS, including AGPE, indicating that the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across all quality measures. When child/family level characteristics are added to the model (Patient Model), the total variance explained by differences between providers does not change significantly.

HRSA study

The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to report their needs met for anticipatory guidance at the follow-up assessment than at the baseline assessment. The table below present comparison of percent of children who received care met the quality care criteria between baseline and follow-up survey data.

Table 5c: Anticipatory Guidance & Parent Education Measure by Children's Characteristics
Parent had their needs met on all AGPE topics

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	38.9% (216)	45.2% (146)	0.08
10-18 months	48.5% (208)	45.7% (150)	0.46
19-48 months	55.0% (193)	65.9% (147)	0.01
Race			
Hispanic	46.0% (46)	47.8% (46)	0.86
White	46.2% (475)	51.9% (372)	0.02
Asian	35.7% (10)	52.9% (9)	0.35
Multiple or other	62.5% (15)	33.3% (6)	0.12
Insurance type			
Private or private and public	46.4% (502)	49.9% (339)	0.15
Public only (includes Medicaid,	44.7% (85)	54.9% (89)	0.07

Medicare, CHIP, and Military)			
Other insurance type	(3)	(1)	-
Uninsured	50.0% (6)	(4)	-
At risk of developmental delay			
Low/no risk	47.5% (487)	52.1% (285)	0.09
High/moderate risk	40.7% (114)	44.4% (76)	0.49

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

Significant gaps and unexplained variations remain in the quality of developmental services for young children. More importantly, parents report unmet informational needs about anticipatory guidance and parental education topics related to behaviors that have an impact on development and behavior.
The probability of receiving anticipatory guidance and parental education varies nearly as much across children seeing the same provider as across providers. The AGPE quality measure assessed here provides a relatively comprehensive picture of performance in the area of preventive and developmental services for young children.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note*: This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Table 2: Proportion of all children in the study who met criteria for receiving quality developmental services across six components of care and ranges in proportion across providers and offices. (SD=Standard Deviation)

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The AGPE items were developed based on several rounds of cognitive interviews with parents to ensure quality of responses appropriate to the questions and minimize missing responses.

Online PHDS: Rate of survey completion was calculated based on survey start and complete dates for each respondent. According to the quality measure scoring protocol, if a parent answered less than half of the items in the quality measure, their score is considered to be missing. This does not include items that should have been appropriately skipped. Missing responses are not given a valid score and are not included in the calculation of the quality measure.

KPNW Study: Of the 5,755 sampled children, 2,173 surveys were returned (37.8%). For these children, the provider the parent identified and the provider to which the child was assigned by the health plan were the same 97.3% of the time. A 95% response rate was obtained for the provider survey.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Online PHDS data show that overall 2.6% of parents who started the survey did not complete the survey. Table 2b7.2a presents the frequency of missing values for Anticipatory Guidance and Parental Education composite measure across all providers and for 5 providers with highest number of surveys.

Table 2b7.2a. The frequency of missing values for Anticipatory Guidance and Parental Education measure, overall and top 5 providers

Quality measures	Overall	Provider ID				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Anticipatory guidance and parental education	1.9 (107)	2.1 (2)	0	3.3 (3)	0	0

KPNW study: Children whose parents responded were not different from those who did not respond in terms of their gender and insurance type. The responding population were slightly less likely to be in the 19 to 48 month age group (55.8% sampled, 51.5% responding) and were somewhat more likely to have had more than one well-visit in the past (67.5% sampled, 74.7% responding).

Table 2b7.2b. Sociodemographic Characteristics of KPNW Starting and Responding Sample

Characteristic	Proportion of Starting Sample (N=5755)	Proportion Respondents as of (N=2162)
Gender of Child[NS]		
Male child	52.7	53.7
Female child	47.3	46.3
Age of the Child[S]		
Child age 3-9 months	19.4	21.8
Child age 10-18 months	24.9	26.7
Child age 19-48 months	55.8	51.5
Type of Insurance[NS]		
Private	98.6	98.5
Public	1.4	1.5
Child's Health Care Utilization		
Number of well-child visits [S]		
1 Well-Child Visit	32.5	25.3
2 or More Well-Child Visits	67.5	74.7
Number of emergency room/urgent care visits		
0 ER/urgent care visits	49.8	51.0
1 ER/urgent care visit	26.2	25.8
2 or more ER/urgent care visit	24.0	23.2
Number of overnight hospital stays [NS]		
0 overnight hospital stays	96.6	96.9
1 or more overnight hospital stays	3.4	3.1

[S]Denotes variables for which statistically significant variation exists between the starting and responding sample for the target child or respondent characteristic.
[NS]No significant variation exists between the starting and responding sample for the target child or respondent characteristic.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Information about non-respondent is not available to compare with those who responded the survey because online PHDS is publicly available tool. However, the low rate of incomplete survey (2.6%) suggests that the measure was acceptable to respondents. Overall, the quality measure had less than 2% of missing cases, ranging 0-3.3% across the top 5 providers with highest number of surveys. Few overall missing values suggest that the measure level results are unlikely to be biased by non-response to the survey questions.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Other
If other: Data are generated by parents completing the Promoting Healthy Development Survey (PHDS).

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.
Patient/family reported information (may be electronic or paper)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).
Data are parent-report using the CAHMI developed Promoting Healthy Development Survey (PHDS). CAHMI captures the data at the provider level through a process described above and in the Evidence Form, Figure 1.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.
During 2012-2016 we experienced some operational delays. in the past several years. In 2012, the provider feedback reports were not automated. When providers wanted a summary report, CAHMI had to manually create them. This was excessively time consuming and CAHMI did not have resources to continue the manual generation of the reports. We sought and received funding to automate the reports. Some difficulties with contractors and staff change-over caused major delays in the project. Then, CAHMI moved from the Oregon Health & Sciences University to Johns Hopkins University School of Public Health in 2014, and it was necessary to upgrade the CAHMI servers. No technical support was available for the transition which caused further delays. Additionally, the PHDS was originally developed in 2001; thus much of the coding and back-end technology for this tool was antiquated and ceased to function after the move. Consequently, and as a result of new improved technology, we have had to redesign the two PHDS related websites - the PHDS toolkit and the parent survey -- as well as the CAHMI PHDS database. Lack of funding caused delays. However, we anticipate launching the new PHDS in February 2017.

Time and cost of data collection are low: provider registration takes about 10 minutes and the parent survey takes about 15-20 minutes to complete. To date, implementation has been limited by lack of funding and resources for outreach, communication

and technical support. Our experience in the development and evaluation of the PHDS demonstrated a clear and compelling need to work closely with providers to overcome the many myths that both parents and providers have about patient-engagement quality improvement tools. For the PHDS to be adopted by providers, it is essential to demonstrate, for example, that tool adds value for both the parent and provider, that it fits into and typically improves work flow in the office; improves parent-provider communication, and most important, improve the quality and delivery of nationally recommended services for children. This can only really be accomplished by collaboration and partnership with providers.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Professional Certification or Recognition Program Quality Improvement (Internal to the specific organization)	

4a.1. **For each CURRENT use, checked above (update for maintenance of endorsement), provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

NA

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
The PHDS toolkit (www.phstoolkit.org) and the parent-reported PHDS (www.wellvisitsurvey.org) were used by 68 uniquely identified providers across the country through 2013. We are happy to provide a list of these providers to NQF if desired. In 2014, CAHMI moved from the Oregon Health & Sciences University, Portland OR to the Johns Hopkins University, Baltimore, MD. As a result of the move, and because both server and database technologies had rapidly evolved and improved over the past few years, it was necessary to upgrade our servers, which in turn caused some technical issues with the links between the provider toolkit, the PHDS, and the CAHMI PHDS database. Additionally, the PHDS was originally used to compare providers within a practice as well as between practices within a health system. The anticipated use of the Online PHDS is intended to provide feedback only for individual providers but not between providers. The combination of these factors led to a decision to upgrade and redesign the PHDS toolkit, PHDS database and Parent Survey. (The PHDS parent survey itself, however, remains fully operational, although use has been nominal from 2014-present, and can be accessed at www.wellvisitsurvey.org.) The redesign required additional time, IT and CAHMI staff resources and delays were incurred during 2014-2015. However, we are now in the process of finalizing the PHDS Toolkit and database redesign, which is anticipated to be completed in February 2017.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

The redesigned PHDS system (registration toolkit, parent survey tool and PHDS database) is anticipated to be completed and fully functional by February 2017. We have a communication and outreach plan to promote the PHDS as part of the CAHMI Cycle of Engagement (see Attachment A-5), which includes the CAHMI Well Visit Planner (www.wellvisitplanner.org) -- a free parent engagement tool that helps prepare parents for the upcoming well child visit – and the post-visit PHDS which assesses whether the parent received services in alignment with national guidelines as well as family centered care. We have been promoting the Cycle of Engagement in national meetings (AMCHP, PAS, APHA, AcademyHealth ARM, National Child Heath Policy Meeting, and more) over the past several years. We presented the Cycle of Engagement at the CMS Quality Meeting December 13, 2016 and have further plans to unveil the redesigned version at meetings in 2017. The WVP and PHDS have also been endorsed tools that meet requirements for Bright Futures implementation.

We have received substantial interest in the CAHMI parent-engagement tools (both the WVP and the PHDS) from and are in extensive conversations with a number of organizations and agencies including health systems, payers, provider organizations – (CMS/Medicaid, Title V, Head Start, Kaiser Permanente and others); professional associations such as the American Academy of Pediatrics, Bright Futures, National Medicaid Medical Directors, the Children's Hospital Association (CHA), AcademyHealth, Association of Maternal and Child Health Programs (AMCHP), CityMatCH, National Initiative for Children's Healthcare Quality (NICHQ), Autism Speaks, Prevent Child Abuse America; National Prevention Information Network (NIPN); national community-based programs and organizations; philanthropic funders; software platform and electronic medical records systems developers and family organizations. We are in the process of securing funding for Cycle of Engagement EMR integration and implementation projects in partnership with or from a number of interested parties. Further, we are finalizing our application to the American Board of Pediatrics to have the Online PHDS certified as a web-based Maintenance of Certification (MOC) (Part 4) quality improvement (QI) tool for pediatricians. ABP has expressed significant interest in the PHDS and provided some initial funding for the redesign efforts.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

In a 2010-2012 study of a large pediatric practice in Oregon (n=551 providers), anticipatory guidance and parental education for physical care increased from 70.3% (n=379) at baseline (2010) to 77.6.(n=197, 2011-12, AOR:1.67, CI:1.11-2.50) post implementation of the CAHMI Well Visit Planner - a family engagement tool to assist parents in planning for their well child visit. This represents a 10.4% increase and was statistically significant at the 95% confidence level. The PHDS, which contains the Anticipatory Guidance and Parent Education measure, was used as the evaluation tool.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**

There were no unintended or unexpected consequences that we are aware of.

4c.2. **Please explain any unexpected benefits from implementation of this measure.**

There were no unexpected benefits that we are aware of.

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
Extensive qualitative interviews with providers and parents have been conducted and previously reported (See Attachment_AGPE Evidence Report)

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
Key informant interviews and focus groups with parents and providers were held throughout the testing and evaluation period. We obtained baseline and post-implementation information from providers and post-implementation information from parents. It was necessary to work closely with practices to demonstrate value of the family engagement tools (Well Visit Planner and PHDS) as well as to modify the process to fit individual practice office culture and work flow. A significant amount of provider and staff education was needed to overcome fears and myths that the tool would add to, not help, time management and that parents would not want to participate. This was accomplished by continued and persistent relationship building, spending much time in the office setting with the staff and providers and holding frequent Q&A sessions as the process unfolded.

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.
Through key informant interviews and focus groups with parents and providers.

4d2.2. **Summarize the feedback obtained from those being measured.**
The Anticipatory Guidance and Parental Education measure is seen as an excellent way by which practices can improve the quality of the visit. In particular this matters a great deal to the providers who are being financially incentivized for family-centered care outcomes.

4d2.3. **Summarize the feedback obtained from other users**
For the most part, parents appreciated being asked about their experience with their well child visits and used it as a way to provide confidential feedback to the providers.

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
The feedback was helpful for future implementation efforts of CAHMI's family engagement tools. The feedback, however, did not result in any changes to the PHDS tool itself.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization of Related Measures
The measure specifications are harmonized with related measures;
OR

The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications harmonized to the extent possible?
No

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
NA

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
NQF #0011 - the PHDS (Promoting Healthy Development Survey) - was endorsed by NQF on October 4, 2012. The PHDS contains the AGPE measure. Neither the questions nor the scoring of the questions have changed since the PHDS was endorsed. It is not actually a competing measure; rather, the AGPE measure is embedded in the PHDS tool.

Please note: The PHDS endorsement (#0011) can be found on the NQF measures website but does not appear to be found in the NQF directory in Question 5 above. Hence, we were forced to enter a "no" to Q5 in order to submit this application.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Attachment_A_Supplemental_Materials_Final.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Child and Adolescent Health Measurement Initiative
Co.2 **Point of Contact:** Christina, Bethell, cbethell@cahmi.edu, 443-287-5092-
Co.3 **Measure Developer if different from Measure Steward:** Child and Adolescent Health Measurement Initiative
Co.4 **Point of Contact:** Christina, Bethell, cbethell@cahmi.edu, 443-287-5092-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
National Advisors for Patient Centered Quality Improvement of Well-Child Care:
Betsy Anderson, Family Voices
David Bergman, Stanford University
Dimitri Christakis, University of Washington
Paula Duncan, University of Vermont
Cynthia Minkovitz, Johns Hopkins School of Public Health
Amy Perritti, American Academy of Pediatrics
Ed Schor, The Commonwealth Fund
Judy Shaw, University of Vermont
Sara Slovin, Johns Hopkins Medicine

Measure Developer/Steward Updates and Ongoing Maintenance Ad.2 **Year the measure was first released:** 2002 Ad.3 **Month and Year of most recent revision:** 01, 2017 Ad.4 **What is your frequency for review/update of this measure?** 3 years Ad.5 **When is the next scheduled review/update for this measure?** 01
Ad.6 **Copyright statement:** None Ad.7 **Disclaimers:** None
Ad.8 **Additional Information/Comments:** None

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3220
Corresponding Measures:
Measure Title: Ask About Parental Concerns
Measure Steward: Child and Adolescent Health Measurement Initiative
Brief Description of Measure: This measure is used to assess the proportion of children whose parents were asked by their child´s health care provider if they have concerns about their child´s learning, development and behavior.
Developer Rationale: Patient-centered care ensures that patients are asked questions about their concerns and that their concerns are addressed. More specifically, parental concerns about their child´s learning, development and behavior have been shown to be reliable and valid indications of a child´s potential risk for developmental, behavioral, or social delays. Asking about parental concerns is a core component of the guidelines set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau. A core component of developmental surveillance, as recommended by the AAP, is to ask the parent at every well-child visit if they have concerns about their child´s learning, development and behavior. Assessment of whether providers asked about parental concerns often cannot be obtained through medical records and administrative data. This information is most valid when collected from the parent regarding their experience of care. Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child´s health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews which are not accurate for the specific level of information obtained in the Promoting Healthy Development Survey (PHDS).

Numerator Statement: The numerator measures the number of parents who had a well child visit within the last 12 months and who indicated that they were asked about their concerns about their child
Denominator Statement: Children age 3 months to 48 months who received a well-child visit in the last 12 months and whose parents responded to the items Ask About Parental Concerns (see Attachment A-2, page 14) on the Promoting Healthy Development Survey (PHDS: www.wellvisitsurvey.org)
Denominator Exclusions: Missing data for the Ask About Parental Concerns questions are excluded from analysis

Measure Type: Outcome: PRO
Data Source: Other
Level of Analysis: Clinician : Individual

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence

asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary

- This is a Patient-Reported Outcome-Based Performance Measure (PRO-PM) derived from the responses to three questions on the Promoting Healthy Development Survey (complete survey starts on page 20 of the Appendix).
- The developer provided a logic model in both graphic and narrative: (1) the parent and child attend a well child visit with their provider; (2) the provider subsequently sends a survey -- the Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org), which includes 1-2 questions on Ask About Parental Concerns (see Attachment A-2, page 14) for the parent to complete; (3) when at least ten surveys have been completed, the provider receives a feedback report on parents' experiences of the visit and the extent to which they felt their concerns were asked about via the CAHMI PHDS Toolkit website (www.phdstoolkit.org); (4) the provider reviews the report and then can engage in a Plan-Do-Study Act (PDSA) quality improvement process to improve their AGPE score. The developer also notes that, recommended developmental services, as set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau, include assessment on alcohol and drug use; presence of guns; family violence; and other safety issues in the family.
- The developer also provides the following support:
    - Patient-centered care ensures that patients are asked questions about their concerns and that their concerns are addressed.
    - Parental concerns about their child's learning, development and behavior have been shown to be reliable and valid indications of a child's potential risk for developmental, behavioral, or social delays.
    - Asking about parental concerns is a core component of the Bright Futures guidelines set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau. The AAP recommends that providers ask the parent at every well-child visit if they have concerns about their child's learning, development and behavior.
- In the Performance Gap section, the developer reports that a HRSA study "found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to report their needs met for Ask About Parent Concerns at the follow-up assessment than at the baseline assessment; and parents were more likely to be asked about one or more psychosocial (family assessment) topics, including asking about their concerns, at follow-up." The results are included in the testing attachment.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Guidance from the Evidence Algorithm: Patient-reported outcome (Box 1)→ Relationship between PRO-PM and provider action (Box 2)→ Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

The developer reports the following:

- An analysis of the top 5 (by number of completed surveys) individual providers had performance scores ranging from 64.9% to 92.3%.
- The Kaiser Permanente Northwest (KPNW) study found just over half of parents reported being asked about whether they had any concerns about their child's developmental and/or behavior – 53.3%. Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant (p=0.002)

Disparities

- The developer reports that the online PHDS showed variation according to a child's age; race/ethnicity (Hispanic=66.3%, white=72.8%, black=63.4%, Asian=65.1%, Other/multi race=60.7%); level of risk for developmental, behavioral, or social delays, respondent education level, and children's special health care needs status. The developer notes that "Children of lower educated mothers are less likely than those with more educated mothers to have high."
- After controlling for other child and family demographic and health factors and provider characteristics, the KPNW study found the likelihood (or adjusted odds ratio-AOR) that a child met quality measure criteria differed significantly according to: (1) child's age (less than 9 months=44.2%, 10-18 months=53.8%, 19-49 months=56.8%) and (2) child's birth order (first born=50.0%, not first born=56.2%).

Questions for the Committee:

o *Is there a gap in care that warrants a national performance measure?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

- This is an outcome measure. The measure needs to clarify not just concerns but specifically about learning, development, and behavior. The reference to both Bright Futures and HRSA were noted. The ages of the children need to be consistent in the numerator and denominator then agree that the measure passes.
- I would argue that asking about concerns is a somewhat tenuous outcome measure (as opposed to, say, parental satisfaction, early detection of developmental delays, etc). But I am willing to say there is a sufficient causal path for this measure
- There is broad professional consensus, and recommendation of the leading profession organization on the importance of asking parents if they have concerns about their child. The authors cite the evidence underlying the new Bright Futures guidelines as well as other sources that parent concerns are a good indication that problem exist. There is value, in any case, of a patient centered approach to care that values the concerns of the patient.
- This is a PRO-PM which establishes a relationship between the measure outcome and a healthcare action. The developers provide several pieces of evidence to support the rational including the AAP Bright Futures guidelines.
- It is unclear how the top 5 providers were determined. There is an incomplete and also inaccurate statement "The developer notes that "Children of lower educated mothers are less likely than those with more educated mothers to have high.'" Rating: moderate.
- There certainly appears to be a gap, which is influenced by a number of factors.
- There is adequate data presented that performance varies, and is less than is expected in high quality care. The data presented on disparities by race/ethnicity lend further support to the existence of a performance gap.
- Yes, a performance gap is provided and described. The developer sites an analysis of individual providers as well as the KPNW study, (showing statistical significance) demonstrating a gap in a core component of the Bright Future guidelines. Disparities are demonstrated in relationship to a child's age, race/ethnicity, level of risk for developmental delays, respondent education level, ect.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Other – patient/family reported survey

Specifications:

- Level of analysis: Clinician – individual
- Interpretation of score: Better quality = Higher score
- Patient Reported Outcome Based Performance Measure (PRO-PM)

- Numerator: *The numerator measures the number of parents who had a well child visit within the last 12 months and who indicated that they were asked about their concerns about their child*
- Denominator: *Children age 3 months to 48 months who received a well-child visit in the last 12 months and whose parents responded to the items Ask About Parental Concerns (see Attachment A-2, page 14) on the Promoting Healthy Development Survey* [Questions are on pg 32 of Appendix A]
- Exclusions: The developer states that "Missing data for the Ask About Parental Concerns questions are excluded from analysis." [NQF does not consider this an exclusion as it is defining the population of the measure.]
- The developer includes a calculation algorithm.
- The measure is not risk adjusted or risk stratified, but the developer states that it can be stratified by variables such as child demographics characteristics (e.g., the child´s age, race); child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or parent health characteristics, if large enough data sets are available.
- The measure does not use sampling.
- This measure relies on a set of questions within the Promoting Healthy Development Survey (pg 33 of the Appendix). This online survey is initiated by the provider who sends it to a parent after a well-child visit. Providers must have a minimum of 10 surveys to generate a report to maintain parent confidentiality.

Questions for the Committee:

- *Are all the data elements (question items) clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

NQF Note: Both measure score and data element reliability testing are required for PRO-PMs.

Method(s) of reliability testing

- The developer used data from the online Promoting Healthy Development Survey (PHDS), a KPNW study, and a HRSA evaluation study that was testing "three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young children in the context of discussions between pediatric clinicians and parents during well-child visits"; the HRSA study used 5 tools including the PHDS.
- The developer did not report data element (item) level reliability testing. The developer noted that "Psychometric item-level reliability testing is not appropriate for AAPC because this measure is not a scale measure. The survey items require "Yes", "No" responses and there is a legitimate skip pattern for the follow-up question. See the attached AAPC Data Dictionary." [*NQF Note: NQF agrees that many of the common psychometric reliability analyses are not appropropriate because this is not a scale measure, however, also notes there are other methods to test item-level reliability.]*
- For score-level testing, the developer used ANOVA to calculate the intra-class correlation (ICC), which assesses the difference between measured objects, and the inter-unit reliability (IUR), which assesses the provider level (score level) reliability.

Results of reliability testing

- No statistical results are available for item-level reliability testing.

- The IUR reliability coefficient for the measure scale is 0.72, which the developer states is within the recommended threshold (above 0.70).
    - [*NQF note: IUR measures the proportion of the measure variability that is attributable to the between facility variance. A small IUR (near 0) reveals that most of the variation of the measures between facilities is driven by random noise, indicating the measure would not be a good characterization of the differences among facilities, whereas a large IUR (near 1) indicates that most of the variation between facilities is due to the real difference between facilities.*]
- The ICC for Asking about Parental Concerns was 0.78 (77.8% of the difference in providers is actual, not due to chance); values above 0.74 are considered excellent according to literature cited by the developer.

Questions for the Committee:

- *The developer does not provide information on the size of the sample used for the IUR, except stating that providers with 10 or more survey were assessed. Does the Committee wish to discuss sample size with the developer?*
- *Is the Committee concerned about the lack of item-level testing?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm: Precise specifications (Box 1)→ Empirical reliability testing (Box 2) → Score level testing (Box 4) → Appropriate method used (Box 5) → High certainty or confidence that the performance measure scores are reliable (Box 6a)→ Moderate

Note: PRO-PMs *require* element-level testing as well, which was conducted and, judged without score-level testing, would be rated MODERATE, the highest rating this testing is eligible for.

Highest possible rating is HIGH.

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Rationale: NQF guidance requires both data element (item)-level and score-level reliability testing. We recognize this PRO-PM is not a scale, but there are other appropriate ways of testing reliability of single items, as noted in feedback provided to the developer during submission. Strictly speaking, the algorithm yields an INSUFFICIENT rating. Committee members should assess whether they are convinced that the item "*In the last 12 months, did your child's doctor or other health provider (could be a general doctor, a specialist, a pediatrician, a nurse practitioner, a physician assistant, a nurse or any one else your child would see for health care) ask if you have concerns about your child's learning, development or behavior?*" itself is reliable absent the empirical data.

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING
Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

☒ **Face validity**

☒ **Empirical validity testing of the measure score**

Validity testing method:

- To assess the concurrent validity of the quality measure, hypothesized associations among PHDS items and scales were examined. The developer tested a hypothesis that "respondents who indicate that providers talked with them about recommended anticipatory guidance topics are less likely to report being concerned about their child's development in related areas compared with respondents who indicate that providers did not talk with them although they wished they had done so."
- Pearson correlation coefficients were calculated across all scale measures to test hypotheses about expected relationships among the PHDS quality measures and to assess the degree to which each of the PHDS quality measures provide unique information. The developer notes that "We expect a moderate or strong correlation between the family assessment scale measures (>0.30) and inter-scale correlation coefficients to be less than 0.80."
- For face validity, the PHDS items were tested using focus groups, in-depth cognitive interviews, a literature review, and an advisory board of expert stakeholders.

Validity testing results:

The developer reports the following ranges of results:

- The concurrent validity testing results demonstrated "showed that significantly fewer parents reported concerns about their child's behavior if they also reported that their child's doctor or other health care providers talked with them about the kinds of behaviors they might expect to see in their child (46.7% "yes, talked" vs 65.5% "no, wish", P < .000; OR: 0.46 95% CI: 0.29 – 0.72)." Similar results were shown regarding parental concerns about how children talk and make speech sounds.
- The developer provides a table of Pearson Correlation Coefficients, which assesses whether the measures are examining different topics. The results suggest, according to the developer, that the measures are not redundant.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

N/A

Questions for the Committee:

- *Is the lack of exclusions consistent with the evidence?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- The developer does not risk adjust the measure because "we do not expect variation in the quality of care provided for children due to risk factors, e.g. children with special health care needs. The provider's performance should be the same regardless of risk factors."

- However, the measure can be stratified by several demographic or health variables as "Identification of variation in quality measures across subgroups of children helps to highlight aspects of care and population of children for which preventive and developmental services may be most need of improvement."
- The developer reports that many studies have shown differences in access to and quality of care, as well as parent satisfaction. The developer states that "One study found: Overall, 94.0% of parents reported 1 or more unmet needs for a number of aspects of care, including assessing family alcohol use, substance abuse and safety. Uninsured children and children aged 18 to 35 months are disproportionately represented among the 15.3% of children whose parents indicated an unmet need this area of care. There are significant variations in performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors."
- Variations were observed by demographic and socioeconomic factors.

Questions for the Committee:

- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To assess meaningful differences, the developer analyzed the top 5 providers (number of individual surveys completed) in the online PHDS; across 56 providers using KPNW study data; and reports on pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.
- Online PHDS: For the top 5 individual providers with the highest numbers of surveys (n=77 to 94) a range of 64.9%-92.3% of parents of young children reported the health care provider asked parents whether they have concerns about their child's development and/or behavior; the average for all children was 69.8%.
- KPNW Study: 53.3% of children had parents reporting their child's well-child care provider asked them whether they had any concerns about their child's development and/or behavior. Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant, 20.0% to 76.0% (p=0.002). Provider n ranged from 15-153.
    - When the provider was used as the level 2 clustering variable, only 1.1-2.2% of the total variance observed was explained by either measured or unmeasured differences between providers. The developer indicates that this "suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers" and that "the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across the all quality measures" (i.e., providers are inconsistent and going to a different provider may not improve a child's care). However, the HRSA study does demonstrate that providers can improve their performance with an intervention.
- HRSA study: The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to be asked about their concerns regarding their child's development or behavior at follow-up, across all children's characteristics measured. For example, for children ages 3-9 months, baseline performance was 64.6% and follow up performance was 73.7 %.

Question for the Committee:

- *Does this measure identify meaningful differences in quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

The developer reports the following:

- **Online PHDS:** Rate of survey completion was calculated based on survey start and complete dates for each respondent. According to the quality measure scoring protocol, if a parent answered less than half of the items

in the AFAUSAS measure, their score is considered to be missing. This does not include items that should have been appropriately skipped. Missing responses are not given a valid score and are not included in the calculation of the quality measure.

- **Online PHDS** data show that 2.6% of parents who started the survey did not complete the survey. The frequency of missing items for the Asking about Parental Concerns measure overall was 2.4%, ranging from 0-4.%.
- **KPNW Study:** Of the 5,755 sampled children, 2,173 surveys were returned (37.8%). For these children, the provider the parent identified and the provider to which the child was assigned by the health plan were the same 97.3% of the time. A 95% response rate was obtained for the provider survey.
- The developer notes that responses for the KPNW survey did not differ by gender or insurance type, but did differ by age and by number of previous well visits.
- The developer states information about non-respondents is not available, but "Overall, the quality measure had 2.4% of missing cases, ranging 0-4.3% across the top 5 providers with highest number of surveys. Few overall missing values suggest that the measure level results unlikely to be biased by non-response to the survey questions."

Guidance from the Validity Algorithm: Specifications consistent with evidence (Box 1)→ Threats to validity addressed (Box 2) →Empirical validity testing (Box 3)→ Measure score testing (Box 6)→ Appropriate method (Box 7)→ Moderate certainty or confidence that the performance measure scores are a valid indicator of quality (Box 8b)→ Moderate

The highest possible score is MODERATE.

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
RATIONALE: Missing data are not fully addressed; non-respondent bias not available.

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

- Patient reported information is not as accurate for outcome measures. Again the "concerns" need to be specified. We note that NQF does not consider the missing data as an exclusion.
- I am worried that systematic confounders may well affect the completion of an online survey.
- The logic is clear, though is a bit confusing given the title of the measure. If the provider asks about concerns, and the patient HAS concerns and does not receive specific information to address them, the provider "fails" the measure. So, the measure is not just about "Asking about parental concerns". It is asking about and addressing them.
- The data elements are clearly defined. A calculation algorithm is included. Exclusion criteria needs to be clarified. Implementation is dependent upon the provider initiating a survey, therefore inconsistency in implementation is possible.
- There is a concern regarding determination of causation when "3 different strategies" were used. It is noted that data element (item) is not reported by the developer. It is also noted that sample size is not reported by the developer. Rating: moderate.
- Is light. I would have appreciated seeing item level reliability testing. I am on the low to insufficient level here.
- The developer does not provide item level reliability testing. IUR and ICC are used to assess reliability. It is not completely clear to me that a sample size of 10 is enough to produce a reliable estimate for a given provider. Especially if this is used for accountability. I did not see an analysis that shows how many surveys per provider would really be needed to do so.
- Provider assessment on only 10 surveys does not seem like an adequate sample size to draw conclusions regarding provider performance. Testing was conducted at the sore level but not the data element level. Item-level reliability testing would strengthen the measure. Results of reliability testing are sufficient to identify differences in performance at the score level.
- It is reassuring that there was empirical validity testing of the measure in addition to face validity. There is concern that there is no "n" size and that this should be an actual measurable outcome such as earlier identification, intervention, etc.
- I struggle with the overall validity. It strikes me as a somewhat "feel good" measure. What is the difference in outcomes from talking about expected developmental milestones versus asking about concerns? Is asking about concerns without talking about the normal ok? I think there are likely important SDS factors that influence this measure! The findings of the level 2 clustering were particularly telling and make me question the utility of this construct.

- "The developer relies somewhat on the validity testing in the development of the questions. The scope of the testing is adequate but broader testing would provide more confidence. The online tool is likely used by a highly selected group of clinicians. The KPNW sample is better, but may not be representative of patients or providers in the US.
- There is the concern that because of recall or other issues, parents may answer ""no"" when the question actually had been asked (though maybe not in a way that was salient or memorable. So, this is like many patient reports of care- they are the best reporters of what happened, but may not be perfect. Given this, I am concerned that the survey would allow the parent to provide data about a visit that happened up to 12 months ago. I don't see any analysis of the time between visit and survey completion on the results.
- I also am concerned by the analysis that includes clustering by provider as the second level. The fact that a patient moving from the lowest to highest provider only has marginally increased odds of meeting the metric concerns me. I would have expected this to be more of a stable measure of a provider that would not vary (within provider) by patient characteristics. It would be good to get more information on interpretation by the developer."

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- This is a patient reported outcome. Data are generated by parents completing the CAHMI-developed Promoting Healthy Development Survey (PHDS), which is sent to them by their provider following a well child visit.
- Although the survey has been in use since 2001, there is not currently an automated reporting system for providers. The developer has been working on a new website for the survey that will automatically report data, and expects it to launch in February 2017.
- The developer reports that the provider registration takes about 10 minutes and the parent survey takes about 15-20 minutes.

Questions for the Committee:

- *Is the data collection strategy ready to be put into operational use?*
- *Does the developer have a status update on the new website?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

- n/a
- Oh sure, its possible. Just another incredibly burdensome data collection and reporting process. And the end result will be a checkbox on the EMR--"I asked the parents/guardian/caretaker about concerns"
- The use of the measure requires use of the PHDS. The developers relate some substantial technical barriers to getting this up and running over the past decade or more. The application states that the barriers should be solved by Feb 2017. It would be good to get an update on this.
- Feasibility of data collection would be increased with the launching of the new website discussed by the developer.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No** ☐ **UNCLEAR**

OR

Planned use in an accountability program? ☐ **Yes** ☒ **No**

Accountability program details
No confirmed use for an accountability program, but CAHMI has been in discussion with a number of organizations that are interested in using the measure, including CMS/Medicaid, Title V, and Head Start.

Improvement results
The developer provided the following response: "Based on PHDS feedback results from an evaluation of the WVP conducted in 2011-2012 in Oregon, we found that Asking about parental concerns did not change significantly because the quality on this measure was already high: Baseline assessment (2010) for this measure was 84.3% and post-assessment testing showed about the same, 83.3% (AOR:0.94, CI: 0-58-1.54)."

Unexpected findings (positive or negative) during implementation
The developer was not aware of any unintended consequences.

Potential harms
The developer was unaware of any potential harms.

Vetting of the measure
The developer conducted key informant interviews and focus groups with patients and providers during testing. The developer reports that "The feedback was helpful for future implementation efforts of CAHMI's family engagement tools. The feedback, however, did not result in any changes to the measure itself."

Feedback:
N/A

***Questions for the Committee**:*
- o *How can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 4: Usability and Use

- n/a
- Arg. Another survey. Another task for the patient/parent. Another checkbox to the EMR. Really???
- Usability is dependent on the website and program being up and running. The use of this for MOC credit is excellent.
- The measure was not publicly reported. There do not seem to be any unintended consequences.

Criterion 5: Related and Competing Measures

Related or competing measures
This measure is part of a set of five based on the PHD survey:
- 3219: Anticipatory Guidance and Parental Education
- 3220: Ask About Parental Concerns
- 3221: Family Centered Care
- 3222: Assessment of Family Alcohol Use, Substance Abuse and Safety
- 3223: Assessment of Family Psychosocial Screening

Harmonization
N/A

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

RATIONALE IF NOT ELIGIBLE:

The measure is not eligible for Endorsement+ because it has not been vetted.

Pre-meeting public and member comments

-

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2974
Measure Title: Ask About and Address Parental Concerns
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: 1/13/2017

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome
☒ Health outcome: Click here to name the health outcome
☒Patient-reported outcome (PRO): Ask About Parental Concerns

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Figure 1 (attached) shows the logic model by which the Ask About Parental Concerns quality measure is obtained and improved. Simply said: (1) the parent and child attend a well child visit with their provider; (2) the provider subsequently sends a survey -- the Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org), which includes 1-2 questions on Ask About Parental Concerns (see Attachment A-2, page 14) for the parent to complete; (3) when at least ten surveys have been completed, the provider receives a feedback report on parents' experiences of the visit and the extent to which they felt their concerns were asked about via the CAHMI PHDS Toolkit website (www.phdstoolkit.org); (4) the provider reviews the report and then can engage in a *Plan-Do-Study Act* (PDSA) quality improvement process to improve their AGPE score. THE PDSA cycle involves reviewing the baseline data; developing and implementing a plan of action to improve the score; obtaining further data from the parent; and comparing the first set of results with the second. The full process is repeated until providers are satisfied with their improved scores. We are currently applying for this process to be approved by the American Board of Pediatrics (ABP) for maintenance of certification (MOC, Part 4) credit. The provider must complete three PDSA cycles. Each time point must have at least 25 completed surveys and there must be at least 8 weeks between time periods.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

Patient-centered care ensures that patients are asked questions about their concerns and that their concerns are addressed. More specifically, parental concerns about their child's learning, development and behavior have been shown to be reliable and valid indications of a child's potential risk for developmental, behavioral, or social delays. Asking about parental concerns is a core component of the guidelines set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau (1). A core component of developmental surveillance, as recommended by the AAP, is to ask the parent at every well-child visit if they have concerns about their child's learning, development and behavior. Assessment of whether providers asked about parental concerns often cannot be obtained through medical records and administrative data. This information is most valid when collected from the parent regarding their experience of care. Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child's health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews which are not accurate for the specific level of information obtained in the Promoting Healthy Development Survey (PHDS).

The process outlined in the logic model in 1a.12 allows health care providers to better understand the extent to which their patients experience "quality care" – in this case, the extent to which parents felt they were asked about their parental concerns. It also allows providers to engage in quality improvement activities to improve their parent-reported quality scores for "Ask About Parental Concerns" by using several Plan-Do-Study Act (PDSA) cycles, as described above.

1. *Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents,* 3rd Edition,

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

Figure_1_Ask_About_Parental_Concerns_Logic_Model.docx,CAHMI_Ask_About_Parental_Concerns_evidence_attachment_revised_02_02_17.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

No

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

IF a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)

IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.

Patient-centered care ensures that patients are asked questions about their concerns and that their concerns are addressed. More specifically, parental concerns about their child´s learning, development and behavior have been shown to be reliable and valid indications of a child´s potential risk for developmental, behavioral, or social delays. Asking about parental concerns is a core component of the guidelines set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau. A core component of developmental surveillance, as recommended by the AAP, is to ask the parent at every well-child visit if they have concerns about their child´s learning, development and behavior. Assessment of whether providers asked about parental concerns often cannot be obtained through medical records and administrative data. This information is most valid when collected from the parent regarding their experience of care. Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child´s health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews which are not accurate for the specific level of information obtained in the Promoting Healthy Development Survey (PHDS).

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis**. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

See also Testing Form-Ask About Parental Concerns.

DATA SOURCES:

Differences in the quality measure scores across providers is demonstrated for (1) 5 top individual providers with the highest number of surveys using Online PHDS data; (2) across 56 providers using KPNW study data; and (3) pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.

Online PHDS: The performance scale for the quality measure was calculated using the scoring methods described in the Attachment . Individual provider level differences in performance were illustrated by the proportion of children meeting the quality of care criteria across 5 top providers with the highest number of completed surveys after their well-child visit.

KPNW Study: The significance of differences observed in the proportion of children meeting criteria for the quality measure across pediatric providers (n=56) was evaluated using t-tests. The relative spread in the quality measure score across providers was assessed using the coefficient of variation statistics (standard deviation across providers multiplied by 100%). Multi-level regression models were conducted using the pediatric provider as the level 2 clustering variable, in order to assess the degree to which the probability that a child meets criteria on each quality measure is explained by differences between providers (called the "clustering effect"). In implementing this multi-level regression method (Empty Model), the presence of a significant clustering effect by pediatric providers was estimated prior to accounting for the child and family characteristics associated with each provider. Second, variables related to the child and family characteristics (child's age, gender, race/ethnicity, birth order, developmental and behavioral delay risk status; parent education and risk for depression) were added to the Empty Model to assess how much of the provider clustering effect observed remains after accounting for these characteristics (called the "Patient Model").

HRSA study: Quantitative data results for the baseline (2010) and follow-up (2011-12) study of the intervention sites using the HRSA Evaluation Study data were conducted using basic descriptive statistics to describe each sample and applying chi-square test of statistical significance to assess differences in the quality measure for the baseline and follow-up samples.

PERFORMANCE DATA

Online PHDS: Table 1b.2a present the proportion of children whose care met for the quality measure across 5 providers. The proportion of parents who reported that the health care provider asked parents whether they have concerns about their child's development and/or behavior ranged 64.9%-76.4%. Only 6%-45% of parents of young children reported that their child's pediatric clinician discussed psychosocial topics such as parent emotional well-being and partner support in parenting.

Table 1b.2a: Proportion of Children Meeting Measure Criteria, Top 5 individual providers with highest number of surveys

Characteristics (n=5355)	All Children				
	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
	1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
Asking parents whether they have concerns about their child's development and/or behavior.	69.8%	76.4%	70.3%	69.0%	92.3% 64.9%

KPNW Study: In this study, a little over one-half of children had parents who reported that their child's well-child care provider asked them whether they had any concerns about their child's development and/or behavior (53.3%). Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant (p=0.002)

Table 1b.2b: Proportion of all children in the study who met criteria for receiving quality services and ranges in proportion across providers. (SD=Standard Deviation)

Developmental Services Quality Measures (n = 2173)	Proportion of All Children Meeting Measure Criteria	Range in the Proportion of Children Meeting Measure Criteria Across 51 Pediatric Providers	Relative Variation (COV) in Measure Scores Across Pediatric Providers
Asking parents whether they have concerns about their child's development and/or behavior.	53.3%	20.0% to 76.0% SD: 11%; (p = 0.002)	20.6%

Only providers with n=15 or more PHDS responses are included in the provider level analysis. Provider level n ranges from 15 to 153.

Multi-level analysis: For the Empty Model that used the provider as the level 2 clustering variable, only 1.1% to 2.2% of the total variance observed in whether children met criteria for each of the all quality measures was explained by either measured or unmeasured differences between the providers that they see. This suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers. These findings translate into a 1.19 to 1.29 median odds ratio across all quality measures, including Ask About Parent Concerns, indicating that the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across all quality

measures. When child/family level characteristics are added to the model (Patient Model), the total variance explained by differences between providers does not change significantly.

HRSA study
The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to report their needs met for Ask About Parent Concerns at the follow-up assessment than at the baseline assessment; and parents were more likely to be asked about one or more psychosocial (family assessment) topics, including asking about their concerns, at follow-up. The tables below present comparison of percent of children who received care met the quality care criteria between baseline and follow-up survey data for each measure and overall composite comprehensive care measure.

Table 1b.2c. Asking about Parent's Concerns about Development Measure, by Children's Characteristics
Parent was asked if they had concerns about their child's development

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	64.6% (357)	73.7% (235)	0.01
10-18 months	78.6% (319)	76.6% (246)	0.59
19-48 months	80.3% (282)	82.3% (181)	0.59
Race			
Hispanic	72.0% (72)	84.1% (37)	0.14
White	72.5% (745)	76.4% (542)	0.07
Asian	75.0% (21)	70.6% (12)	0.74
Multiple or other	87.5% (21)	77.8% (14)	0.44
Insurance type			
Private or private and public	72.4% (784)	75.5% (509)	0.16
Public only (includes Medicaid, Medicare, CHIP, and Military)	78.4% (149)	82.2% (129)	0.42
Other insurance type	71.4% (5)	(3)	-
Uninsured	50.0% (6)	100% (7)	0.04
At risk of developmental delay			
Low/no risk	72.2% (741)	76.3% (411)	0.09
High/moderate risk	76.2% (214)	82.4% (140)	0.13

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
NA

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*
See also Testing Form_Ask About Parental Concerns
DATA SOURCES
We used the following data sources for testing of the quality measure:
(1) Online Promoting Healthy Development Survey (PHDS) – data collected through an online, publicly available tool (Promoting Healthy Development Survey-PHDS). Parents who had a well-child care visit in the last 12 months can complete the PHDS. Providers initiate the survey. (See Evidence Form Figure 1 for visual model of the Online PHDS.)
2) Kaiser Permanente Northwest (KPNW) Study – CAHMI partnered with Kaiser Permanente Northwest in Portland, Oregon. The study aimed to evaluate the level and variations in the quality of preventive and developmental services for young children and assess the contribution of key system, provider and patient factors.
3) HRSA Evaluation Study - The specific goal of this study was to evaluate the feasibility, acceptability and impact of three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young

children in the context of discussions between pediatric clinicians and parents during well-child visits. The evaluation measures used data from 5 different tools/surveys including PHDS. The parent-completed PHDS was administered before and after the intervention to assess changes in the quality of well-child care. The study funded by Health Resources and Services and Administration's (HRSA) Maternal and Child Health Bureau. (Patient Centered Quality Improvement of Well-Child Care, Final Report, Supported by a grant from the Maternal and Child Health Bureau Research Grants Program, Health Resources and Services Administration, R40 MC08959 03-00.)

STUDY POPULATION

Online PHDS: Children age 3-48 months of age whose parents completed the online publicly available PHDS were included in the testing. During 2008-2016, we received 5,670 completed surveys. Of those surveys, 5,355 surveys with provider IDs were used for analyses. Children's socio-demographic and health characteristics varied across the individual providers included in the analysis.

Table 1a: Characteristics of children for whose visited provider ID is available

Characteristics (n=5355)	All Children	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)					
		1029 (n=94)		948 (n=91)		1067 (n=90)	
Age of child							
Under 10 months of age	38.3%	19.1%	49.5%	33.3%	54.4%	24.7%	
10 to 18 months of age	34.7%	39.4%	38.5%	38.9%	29.1%	57.1%	
19-47 months of age	27.0%	41.5%	12.1%	27.8%	16.5%	18.2%	
Race/ethnicity of child							
White, non-Hispanic	53.8%	13.3%	81.0%	20.3%	50.7%	17.3%	
Hispanic	40.8%	81.1%	14.3%	74.7%	40.8%	78.7%	
Other race/ethnicity	5.3%	5.5%	4.7%	7.0%	8.4%	4.0%	
Respondent education level							
Did not complete high school	12.1%	23.6%	0	34.1%	6.4%	15.8%	
Completed high school	88.9%	76.4%	100%	65.9%	93.6%	84.2%	
Children who qualify for Children with Special Health Care Needs (CSHCN) Screener criteria							
CSHCN	10.1%	7.4%	8.8%	10.0%	11.4%	5.2%	
Non-CSHCN	89.9%	92.6%	91.2%	90.0%	88.6%	94.8%	
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	22.7%	-	24.4%	-	28.9%	0%	

-Data is not available due to small sample size

KPNW Study: The population studied was children 3 to 48 months old who live in a metropolitan area in the Pacific Northwest. One randomly selected child per household whose age would be no younger than 3 months of age and no older than 48 months of age at the time that their parents received the survey and had one or more well-child visits were eligible to be sampled. A random sample of 5,755 children were identified. Of the 5,755 sampled children, 2,173 surveys were returned (37.8%).

Table 1b: Characteristics of children for whom survey responses were received, KPNW study, Top 5 individual providers with highest number of surveys

Characteristics (n=2173)	All Children	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)					
		7 (n=80)	53 (n=77)		4 (n=74)	1 (n=67)	43 (n=66)
Age of child							
Under 10 months of age	22.0%	20.0%	19.5%	20.3%	22.4%	21.2%	
10 to 18 months of age	26.6%	25.0%	29.9%	35.1%	22.4%	15.2%	
19-47 months of age	51.4%	55.0%	50.6%	44.6%	55.2%	63.6%	
Gender of child							
Female child	46.2%	48.8%	49.4%	47.3%	41.8%	45.5%	
Male child	53.8%	51.3%	50.6%	52.7%	58.2%	54.5%	
Race/ethnicity of child							
White, non-Hispanic	72.9%	84.8%	77.0%	93.2%	76.9%	62.5%	
Asian, non-Hispanic	7.8%	2.5%	6.8%	1.4%	3.1%	20.3%	

Hispanic	8.9%	6.3%	12.2%	2.7%	10.8%	10.9%						
Other race/ethnicity	10.4%	6.3%	4.1%	2.7%	9.2%	6.3%						
Child is the first born in the family	52.1%	52.5%	40.8%	35.1%	54.5%	52.3%						
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	31.3%	21.5%	24.7%	27.0%	29.7%	26.2%						
Education level of mother												
High school or less	12.7%	20.3%	3.9%	14.9%	16.7%	6.2%						
More than high school	87.3%	79.7%	96.1%	85.1%	83.3%	93.8%						

HRSA Evaluation Study: The study inclusion criteria were used to determine which parents/guardians of children were invited to participate in the interventions and/or evaluation from each participating study site:

- Parent has a well-child visit scheduled at this intervention site for one or more of their children.
- The child is scheduled for their 4-month to 3-year-old well-child visit and, therefore, is between the ages of 4 and 40 months (e.g. 40 month old children could be there for their 3 year well-child visit)
- The parent can read and understand English and is able to complete the intervention and evaluation tools.
- For intervention, the parent was able to access the online version of the Plan My Child's Well-Visit tool and the online evaluation survey.

The analysis includes 551 completed surveys at baseline (2010) and 275 completed surveys at follow-up (2011-12)

Table 1c. Sample description for baseline and follow-up PHDS respondents

	Baseline (n=551)	Follow-up (n=275)
Visit type of child for whom survey was completed		
4, 6 or 9-month	38.9%	36.2%
12, 15 or 18-month	33.7%	41.3%
24 or 36-month	27.4%	22.4%
Birth order of child for whom survey was completed		
First child	42.2%	56.6%
Not first child	57.8%	43.4%
Race/ethnicity		
White, non-Hispanic	80.3%	83.5%
Hispanic	8.4%	6.6%
Other/multiple, non-Hispanic	8.6%	6.6%
Asian, non-Hispanic	2.7%	3.3%
Insurance type		
Private or private and public	90.7%	86.7%
Public only (includes Medicaid, Medicare, CHIP and Military)	7.6%	12.1%
Other	0.7%	0.4%
None	0.9%	0.8%

DISPARITIES

Online PHDS: Variation is observed according to a child's age; race/ethnicity; level of risk for developmental, behavioral, or social delays, respondent education level, and children's special health care needs status. Non-Hispanic white children is most likely to meet scoring criteria. Children of lower educated mothers are less likely than those with more educated mothers to have high.

Table 3a. Asking about parental concerns by child demographics and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	1492	66.1%
9-18 months	1399	73.8%
19-48 months	982	70.4%
p values (Pearson chi-square)	-	<0.001
Gender		
Male	396	64.6%
Female	400	64.0%
p values (Pearson chi-square)	-	0.83

Race/ethnicity						
Hispanic	1282	66.3%				
White non-Hispanic	2052	72.8%				
Black non-Hispanic	64	63.4%				
Asian non-Hispanic	71	65.1%				
Other/Multi race, non-Hispanic	51	60.7%				
p values (Pearson chi-square)	-	<0.0001				
Adult survey responds education level						
Did not complete high school	361	60.0%				
Completed high school or higher education	3363	71.1%				
p values (Pearson chi-square)	<0.0001					
CSHCN status						
Non-CSHCN	3431	69.0%				
CSHCN	431	77.0%				
p values (Pearson chi-square)	-	<0.001				
At risk for developmental delay (online only)						
Low/No risk	1645	73.7%				
High/Moderate risk	544	73.4%				
p values (Pearson chi-square)	-	0.88				

KPNW study: After controlling for other child and family demographic and health factors and provider characteristics, the likelihood (or adjusted odds ratio-AOR) that a child met quality measure criteria differed significantly according to: (1) child's age and (2) child's birth order

Table 3b: Mean number of developmental services care components for which quality care was received and the proportion of children meeting criteria for receiving quality developmental services by characteristics of children and families.

Characteristic of Child or Child's Family	% Meeting Criteria
Child's Age	
Less than 9 mos.	44.2% S
10 to 18 mos.	53.8%
AOR: 1.54	
19 to 49 mos.	56.8%
AOR: 1.67	
Child's Gender	
Male Child	54.1% NS
Female Child	52.3%
Child's Race	
White, Non-Hispanic	54.1% NS
Asian, Non-Hispanic	52.7%
Hispanic	51.3%
Other Race, Multiple Race	48.6%
Birth Order	
Not First Born	56.2% S
First Born	50.0%
AOR: .80	
Child's Risk for Developmental, Behavioral or Social Delays (Using Parent's Evaluation of Developmental Status)	
Low/No Risk	52.9% NS
At Risk	54.7%
Respondent Education	
More than High School	53.6% NS
High School or Less	50.7%
Respondent's Risk for Depression (Using the Kemper Screener)	
No Symptoms of Depression	54.4% NS
Symptoms of Depression	49.0%

NOTE: Adjusted odds ratios (AOR) derived from regression analyses listed in the table are shown only if they are statistically significant. AOR uses the first subgroup of each characteristic as a reference. s differences significant at the p < .05 level of significance; NS differences not significant.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4**
NA

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

De.7. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** CAHMI_Data_Dictionary_Ask_About_Parental_Concerns.pdf

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.
No

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.
NA

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator measures the number of parents who had a well child visit within the last 12 months and who indicated that they were asked about their concerns about their child

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator is the number of survey respondents answering either "Yes and my questions were answered" or "No, but I already had information about this topic and did not need to talk about it anymore." Parents must answer BOTH of the two questions for the data to be included in the analysis. An aggregated 100% positive response is needed to achieve quality for this aspect of care.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
Children age 3 months to 48 months who received a well-child visit in the last 12 months and whose parents responded to the items Ask About Parental Concerns (see Attachment A-2, page 14) on the Promoting Healthy Development Survey (PHDS: www.wellvisitsurvey.org)

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
Children age 3 months to 48 months who received a well-child visit in the last 12 months and whose parents responded to the items Ask About Parental Concerns (see Attachment A-2, page 14) on the Promoting Healthy Development Survey (PHDS: www.wellvisitsurvey.org)

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Missing data for the Ask About Parental Concerns questions are excluded from analysis

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
Parents must answer all questions (1, 2 and if applicable 2a) for their data to be included in the analysis. Missing data are excluded from the analysis. Approximately 2.6% of parents who started the Online PHDS did not complete the survey (range 0.0-4.3% for top 5 providers with highest number of surveys; see Testing form, pages 21-22 for more detailed information on missing data).

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*
Although no stratification is required, the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the findings by possible vulnerability, should any individual provide have sufficient data (parent responses) to do so. Potential variables for stratification include:
(1) Child demographic characteristics (e.g., the child´s age, race);
(2) Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or
(3) Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
No risk adjustment or risk stratification
If other:

S.12. **Type of score:**
Rate/proportion
If other:

S.13. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
The numerator is the sum of survey respondents (parents) answering either "Yes to Questions 1 and 2a (if applicable). The denominator is the sum of all respondents answering the Ask About Parental Concerns questions in the PHDS. A score of 100% aggregated across the items represents quality for asking about and addressing parental concerns.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
NA

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Data are collected using the parent-reported "Promoting Healthy Development Survey" (PHDS) developed by the CAHMI (www.wellvisitsurvey.org). Instructions for survey completion are included with the survey. Ask About and Address Parental Concerns questions are multiple choice (Yes or No for all questions. Question #1 also include "I don´t remember" response category. The PHDS is initiated by the provider who can send it to all parents who have received a well child visit. CAHMI has a website (www.phdstoolkit.org) where providers can register to use for the PHDS. This site assigns each provider a unique URL, which allows for provider identification by CAHMI as well as light branding with the provider´s logo so that it is identifiable by the parent. The PHDS Toolkit website sends an email to the provider with the unique URL link to the survey. The provider then sends the link to the parents asking them with instructions to fill out the survey and provide feedback about the visit. The parent fills out the survey and receives a customized feedback report. The survey data are captured on a secure HIPAA compliant CAHMI server. Through the PHDS Toolkit website, providers can generate a report that aggregate parent data information from the survey. Providers must have a minimum of 10 surveys to generate a report to maintain parent confidentiality. See Evidence Form, Figure 1 for a visual model this process.

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Other

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The Ask About Parental Concerns measure is included as part the CAHMI Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org, see Attachment A-2 page 14). The data are generated by parents filling out the PHDS. The PHDS is based in English. See Evidence Form, Figure 1 for a description visual model of the data collection process.

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Individual

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Clinician Office/Clinic
If other:

S.22. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
NA

2. Validity – See attached Measure Testing Submission Form

CAHMI_NQF_Testing_attachment_Ask_About_Parental_Concerns_020217.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care)

What were the statistical results of the analyses used to select risk factors?

Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

No - This measure is not risk-adjusted

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2974
Measure Title: Ask About Parental Concerns (AAPC)
Date of Submission: 2/2/2017
Type of Measure:

☒ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process	☐ Efficiency
☐ Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record

☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Patient reported data	☒ other: Patient reported data

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

We used the following data sources for testing of the quality measure:

1) Online Promoting Healthy Development Survey (PHDS) – data collected through an online, publicly available tool (Promoting Healthy Development Survey-PHDS). Parents who had a well-child care visit in the last 12 months can complete the PHDS. Providers initiate the survey. (See Evidence Form Figure 1 for the Online PHDS logic model.)
2) Kaiser Permanente Northwest (KPNW) Study – CAHMI partnered with Kaiser Permanente Northwest in Portland, Oregon. The study aimed to evaluate the level and variations in the quality of preventive and developmental services for young children and assess the contribution of key system, provider and patient factors.
3) HRSA Evaluation Study - The specific goal of this study was to evaluate the feasibility, acceptability and impact of three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young children in the context of discussions between pediatric clinicians and parents during well-child visits. The evaluation measures used data from 5 different tools/surveys including PHDS. The parent-completed PHDS was administered before and after the intervention to assess changes in the quality of well-child care. The study funded by Health Resources and Services and Administration's (HRSA) Maternal and Child Health Bureau. (Patient Centered Quality Improvement of Well-Child Care, Final Report, Supported by a grant from the Maternal and Child Health Bureau Research Grants Program, Health Resources and Services Administration, R40 MC08959 03-00.)

1.3. What are the dates of the data used in testing? 2004-2016
Online PHDS: 2008-2016
KPNW Study: 2004-2005
HRSA Evaluation Study: 2010-2012

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Online PHDS: n=5,670 surveys reporting on quality of care provided by 299 individual pediatricians and primary care providers from 88 clinics in 36 states. Participation is a voluntary self-selection process based on knowledge and interest in quality improvement in their practice.

KPNW Study: Provider-level surveys and quality of care assessment were focused on the care provided by 56 individual providers (44 pediatricians, 9 nurse practitioners, 3 physician assistants) in the pediatrics department who were organized into ten geographically distinct offices.

HRSA Evaluation Study: Three pediatric offices in Oregon: 1) a rural site, (4 pediatricians), 2) an urban site (8 pediatricians), and 3) an urban site, (12 pediatricians). All pediatricians in selected clinic and office staff participated in relevant baseline and follow up data collection.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Online PHDS: Children age 3-48 months of age whose parents completed the online publicly available PHDS were included in the testing. During 2008-2016, we received 5,670 completed surveys. Of those surveys, 5,355 surveys with provider IDs were used for analyses. Children's socio-demographic and health characteristics varied across the individual providers included in the analysis.

Table 1.6a: Characteristics of children for whose visited provider ID is available

Characteristics	All Children (n=5355)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
Age of child						
Under 10 months of age	38.3%	19.1%	49.5%	33.3%	54.4%	24.7%
10 to 18 months of age	34.7%	39.4%	38.5%	38.9%	29.1%	57.1%
19-47 months of age	27.0%	41.5%	12.1%	27.8%	16.5%	18.2%
Race/ethnicity of child						
White, non-Hispanic	53.8%	13.3%	81.0%	20.3%	50.7%	17.3%
Hispanic	40.8%	81.1%	14.3%	74.7%	40.8%	78.7%
Other race/ethnicity	5.3%	5.5%	4.7%	7.0%	8.4%	4.0%
Respondent education level						
Did not complete high school	12.1%	23.6%	0	34.1%	6.4%	15.8%
Completed high school	88.9%	76.4%	100%	65.9%	93.6%	84.2%
Children who qualify for Children with Special Health Care Needs (CSHCN) Screener criteria						
CSHCN	10.1%	7.4%	8.8%	10.0%	11.4%	5.2%
Non-CSHCN	89.9%	92.6%	91.2%	90.0%	88.6%	94.8%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	22.7%	-	24.4%	-	28.9%	0%

-Data is not available due to small sample size

KPNW Study: The population studied was children 3 to 48 months old who live in a metropolitan area in the Pacific Northwest. One randomly selected child per household whose age would be no younger than 3 months of age and no older than 48 months of age at the time that their parents received the survey and had one or more well-child visits were eligible to be sampled. A random sample of 5,755 children were identified. Of the 5,755 sampled children, 2,173 surveys were returned (37.8%).

Table 1.6b: Characteristics of children for whom survey responses were received, KPNW study, Top 5 individual providers with highest number of surveys

Characteristics	All Children (n=2173)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		7 (n=80)	53 (n=77)	4 (n=74)	1 (n=67)	43 (n=66)
Age of child						
Under 10 months of age	22.0%	20.0%	19.5%	20.3%	22.4%	21.2%
10 to 18 months of age	26.6%	25.0%	29.9%	35.1%	22.4%	15.2%
19-47 months of age	51.4%	55.0%	50.6%	44.6%	55.2%	63.6%
Gender of child						
Female child	46.2%	48.8%	49.4%	47.3%	41.8%	45.5%
Male child	53.8%	51.3%	50.6%	52.7%	58.2%	54.5%
Race/ethnicity of child						
White, non-Hispanic	72.9%	84.8%	77.0%	93.2%	76.9%	62.5%
Asian, non-Hispanic	7.8%	2.5%	6.8%	1.4%	3.1%	20.3%
Hispanic	8.9%	6.3%	12.2%	2.7%	10.8%	10.9%
Other race/ethnicity	10.4%	6.3%	4.1%	2.7%	9.2%	6.3%
Child is the first born in the family	52.1%	52.5%	40.8%	35.1%	54.5%	52.3%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	31.3%	21.5%	24.7%	27.0%	29.7%	26.2%
Education level of mother						
High school or less	12.7%	20.3%	3.9%	14.9%	16.7%	6.2%
More than high school	87.3%	79.7%	96.1%	85.1%	83.3%	93.8%

HRSA Evaluation Study: The study inclusion criteria were used to determine which parents/guardians of children were invited to participate in the interventions and/or evaluation from each participating study site:

- Parent has a well-child visit scheduled at this intervention site for one or more of their children.
- The child is scheduled for their 4-month to 3-year-old well-child visit and, therefore, is between the ages of 4 and 40 months (e.g. 40 month old children could be there for their 3 year well-child visit)
- The parent can read and understand English and is able to complete the intervention and evaluation tools.
- For intervention, the parent was able to access the online version of the Plan My Child's Well-Visit tool and the online evaluation survey.

The analysis includes 551 completed surveys at baseline (2010) and 275 completed surveys at follow-up (2011-12)

Table 1.6c. Sample description for baseline and follow-up PHDS respondents

	Baseline (n=551)	Follow-up (n=275)
Visit type of child for whom survey was completed		
4, 6 or 9-month	38.9%	36.2%
12, 15 or 18-month	33.7%	41.3%
24 or 36-month	27.4%	22.4%
Birth order of child for whom survey was completed		
First child	42.2%	56.6%
Not first child	57.8%	43.4%
Race/ethnicity		
White, non-Hispanic	80.3%	83.5%
Hispanic	8.4%	6.6%
Other/multiple, non-Hispanic	8.6%	6.6%
Asian, non-Hispanic	2.7%	3.3%
Insurance type		
Private or private and public	90.7%	86.7%
Public only (includes Medicaid, Medicare, CHIP and Military)	7.6%	12.1%
Other	0.7%	0.4%
None	0.9%	0.8%

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Online PHDS and KPNW study data were used for reliability testing and stratification analysis. Validity findings are presented from a peer-reviewed publications and online PHDS data. Performance analysis was conducted using the online PHDS, KPNW study and HRSA Evaluation Study data.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Online PHDS: Child's age, sex, race/ethnicity, and respondent (parent) age, race/ethnicity, and education level. The survey does not have a question asks about family income due to complexity of collecting income data by self-reported survey. However, the online PHDS has items assessing the family's economic situation: How much trouble does the family have paying for a) child's health and medical expenses; b) supplies like formula, food, diapers, clothes and shoes; and c) health care for the parent.
KPNW Study: Child's age, sex, race/ethnicity, and education level of mother
HRSA Study: Child's age, race-ethnicity, and insurance type

_____

2a2. RELIABILITY TESTING
***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Psychometric item-level reliability testing is not appropriate for AAPC because this measure is not a scale measure. The survey items require "Yes", "No" responses and there is a legitimate skip pattern for the follow-up question. See the attached AAPC Data Dictionary.

The primary aim of the quality measure is to detect differences between providers on the quality of care provided to young children. Provider level reliability was assessed by inter-unit reliability (IUR) using analysis of variance. IUR can be interpreted as the fraction of the variation among provider scores that is due to real differences, rather than due to chance. If the IUR is higher, the ability of the measure to discriminate across programs is greater. Scales with reliability coefficients above 0.70 provide adequate precision for use in statistical analysis of unit-level comparisons.[1] As the IUR gets smaller, a larger sample is needed in order to reliably discriminate across programs.

Intra-class correlation (ICC) was calculated using ANOVA, as a ratio of the variance between groups over the total variance. The interpretation of the ICC is as the proportion of relevant variance that is associated with differences among measured objects.[2] Fleiss (1981) and Cicchetti and Sparrow (1981) from the medical group state that ICC range categories are: < 0.40 = poor; 0.40 – 0.59 = fair; 0.60 – 0.74 = good; and > 0.74 = Excellent[3]. Values above about 0.7-0.8 are considered acceptable for applied tests. In the analysis we included providers with 10 or more completed surveys.

1. Nunnally, J. C. Psychometric theory (2nd ed). 1978, New York: McGraw-Hill.
2. McGraw, K. O., & Wong, S. P. Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1996:1(1), 30-46.
3. Cicchetti D.V., and Sparrow, S.S. Developing criteria for establishing the interrater reliability of specific items in a given inventory. American Journal of Mental Deficiency, 1981:86, 127-137.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

No statisticsal results are available for item-level reiliabilty testing.

Inter-unit reliabilty coefficient for the measure scale is within the recommended threshold (0.72), indicating that the measure reliably detects differences between providers. Intraclass correlation coefficient for the Asking about Parental Concerns measure is 0.78, indicating that 77.8% of the variance in the mean of the providers is "true" rather than due to chance.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i*.e., what do the results mean and what are the norms for the test conducted?*)

The quality measure provides a reliable assessment of the provision of nationally recommended well-child care with strong inter-unit reliability coefficient (0.72) and intraclass correlation (0.78).

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

A standard, multistage process was used to ensure validity of the AAPC measure:

- Focus groups and in-depth cognitive interviews were conducted throughout the survey development process;
- A review of literature identified through Medline or during key informant interviews; and,
- Three Advisory Groups comprised of pediatricians, family practitioners, consumer representatives, public health experts, and researchers, regularly reviewed and provided input on the identification of quality measurement topics and the development of the PHDS.

A "gold standard" does not exist for determining the criterion validity of patient-reported AAPC measure. However, to ensure the validity of the AAPC quality measure, we followed rigorous procedures representing best practices within the field to develop the survey questions. To ensure the content validity of measures of parent experiences, we used qualitative methods, including both focus groups and cognitive interviews, to inform development and evaluation of the AAPC questions. Focus groups with families aimed to identify the aspects of health care quality that are important to parents in the area of preventive care for their children, including asking about and addressing parental concerns. In-depth cognitive testing of the survey items was conducted with 15 families representing a range of racial, income and education groups as well as different types of health insurance coverage, age of child, age and sex of parent, and number of children in family. Focus groups and cognitive interviews with 35 health care providers in Vermont and Washington and 20 parents of young children in Vermont were conducted to inform item-reduction, administration specifications, and reporting templates. Survey modifications were made based on findings in order to improve the reliability, validity and cognitive ease of the AAPC items.[1]

To assess the concurrent validity of the AAPC quality measure, we tested a hypothesis that respondents who indicate that providers talked with them about recommended anticipatory guidance topics are less likely to report being concerned about their child's development in related areas compared with respondents who indicate that providers did not talk with them although they wished they had done so.

Pearson correlation coefficients were calculated across all scale measures to test hypotheses about expected relationships among the PHDS quality measures and to assess the degree to which each of the PHDS quality measures provide unique information. We expect a moderate or strong correlation between the AACP measure (>0.30) and inter-scale correlation coefficients to be less than 0.80.

[1]Bethell C, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. Pediatrics. 2001 May;107(5):1084-94.

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Using behavior coding methods, for each item in the AAPC quality measure, instances where the respondent required clarification or did not appropriately answer an item were noted. Also, items where the interviewer had difficulty asking the question without edits to the wording were noted. Data analysis was used to inform item-reduction. Content was revised and refined iteratively with each set of interviews.

Cognitive testing confirmed the readability of the AAPC items for people across a range of educational levels. Parents were uniformly able to complete the self-administered survey in 10-15 minutes. Readability assessments indicated the AAPC items to be written at the 8th-9th grade reading level. Survey design and formatting was finalized with input from a group of experts and family representatives.

Concurrent validity testing showed that significantly fewer parents reported concerns about their child's behavior if they also reported that their child's doctor or other health care providers talked with them about the kinds of behaviors they might expect to see in their child (46.7% "yes, talked" vs 65.5% "no, wish", P < .000; OR: 0.46 95% CI: 0.29 – 0.72). Similarly, parents of children 10 to 48 months old were less likely to report being concerned about how their child talks and makes speech sounds if they indicated that their child's doctor or other health care providers had talked with them about words and phrases used by their child (35.2% "yes, talked" vs 48.5% "no, wish", P < .005; OR: 0.58, 95% CI: 0.37–0.89).

Correlations between the AAPC and other PHDS quality measures were not so high as to suggest redundancy across measures (average correlation: 0.34).

Table 2b2.3. Pearson Correlation Coefficients among PHDS Quality Measures (online PHDS)

Scale Measures	Anticipatory Guidance and Parent Education	Family Centered Care	Ask About Parental Concern	Assessment of smoking, drug and alcohol use and safety in the family	Assessment of family psychosocial well-being
Family Centered Care	.52				
Ask About Parental Concern	.16	.14			
Assessment of smoking, drug and alcohol use and safety in the family	.16	.13	.07		
Assessment of family psychosocial well-being	.19	.16	.09	.54	

Average correlation: 0.34

The two AAPC items have been used in two national surveys of parents—The National Survey on Early Childhood Health and the National Survey of Children's Health. The AAPC quality measure is among the few recognized in the Agency for

Healthcare Research and Quality's Child Health Toolbox and the National Quality Measures Clearinghouse as measures that meet basic criteria for use as standardized indicators of health care quality for children.

1. Bethell C, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. Pediatrics. 2001 May;107(5):1084-94.
2. Christina Bethell, PhD, MPH, MBA; Colleen H. Peck Reuland, MS; Neal Halfon, MD, MPH; Edward L. Schor, Measuring the Quality of Preventive and Developmental Services for Young Children: National Estimates and Patterns of Clinicians' Performance. Pediatrics, 2004, 113(6):1973-83

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i*.e., what do the results mean and what are the norms for the test conducted?*)

The AAPC quality measure provides conceptually valid assessment of the provision of nationally recommended preventive care services for young children. Each of the five composite quality measures of PHDS provides unique information about performance. Regardless of the population group or the aspect of health care assessed, the quality of health care rarely can be represented accurately by either a single composite performance measure or by assessing whether a single recommended service is provided. The measure is used in national surveys and recognized as measures that meet basic criteria for use as standardized indicators of health care quality for children. The measure serves as an important complement to existing quality measures.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Not applicable

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Not applicable

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Not applicable

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☒ **Stratification by** variable number of **risk categories**

☐ **Other,**

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

Not applicable

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

The AAPC quality measure does not require risk adjustment because we do not expect variation in the quality of care provided for children due to risk factors, e.g. children with special health care needs. The performance should be the same regardless of risk factors. The national experts extensively reviewed the risk adjustment requirements during development of the measure items of the PHDS tool and did not recommend risk-adjustment for any of the measures. In addition, during the KPNW study, we assessed whether the probability of receiving guidance, education or screening was higher according to a child's level of need or risk, thereby indicating that providers are customizing care to children. The study found no evidence that providers customize care to children most at risk.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Identification of variation in the AAPC quality measure across subgroups of children helps to highlight aspects of care and population of children for which preventive and developmental services may be most need of improvement. Although no stratification is required (number of surveys for each individual providers may not be sufficient to stratify), the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the quality measures by possible vulnerability:

- Child demographic characteristics (e.g., the child's age, race)
- Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs)
- Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

Based on extensive literature review and expert panel, we identified that child and parent demographics such as age, sex, race-ethnicity, income, insurance, parent behavior, CSHCN screener and follow-up for children at risk can be used for stratification. Several studies have documented differences in access and quality of care provided to children, as well as in parent-reported satisfaction with care.[1-2] One study found: "Overall, 94.0% of parents reported 1 or more unmet needs for parenting guidance, education, and screening by pediatric clinician(s) in 1 or more of the content of care areas evaluated (including asking about and addressing parental concerns). Uninsured children and children aged 18 to 35 months are disproportionately represented among the 15.3% of children whose parents indicated an unmet need in each of the 4 areas of care. There are significant variations in performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors."[3]

The KPNW study assessed child and family characteristics to characterize the child and their family based on the PHDS item responses: child's race/ethnicity, birth order, risk for developmental, behavioral, or social delays using responses to Frances Glascoe's Parents' Evaluation of Developmental Status (PEDS) items included in the PHDS 29 parent's education; and whether he/she is experiencing symptoms of depression using Kathy Kemper's screening items. Adjusted odds ratios were calculated using logistic regression analysis in order to assess differences in the odds of meeting quality measure criteria according to child, family and provider characteristics, after controlling for other variables.

References:

1. Halfon N, Regalado M, Sareen H,Inkelas M, Reuland CH, Glascoe FP, Olson LM. Assessing development in the pediatric office. Pediatrics. 2004 Jun;113(6 Suppl):1926-33.
2. Weech-Maldonado R, Morales LS, Spritzer K, Elliott M, Hays RD. Racial and ethnic differences in parents' assessments of pediatric care in Medicaid managed care. Health Serv Res. 2001 Jul;36(3):575-94.
3. Bethell C, Reuland CH, Halfon N, Schor EL. Measuring the quality of preventive and developmental services for young children: national estimates and patterns of clinicians' performance. Pediatrics. 2004 Jun;113(6 Suppl):1973-83.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

See 2b4.3.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Pearson's chi-squire test was used to compare the prevalence of AAPC quality measure across the stratification characteristics. We preformed logistic regression analysis in order to assess differences in the odds of meeting quality measure criteria according to child, family and provider characteristics, after controlling for other variables.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Not applicable

2b4.9. Results of Risk Stratification Analysis:

Online PHDS: Variation is observed according to a child's age; race/ethnicity; level of risk for developmental, behavioral, or social delays, respondent education level, and children's special health care needs status. Non-Hispanic white children is most likely to meet scoring criteria. Children of lower educated mothers are less likely than those with more educated mothers to have high.

Table 2b4.9a. Asking about parental concerns by child demographics and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	1492	66.1%
9-18 months	1399	73.8%
19-48 months	982	70.4%
p values (Pearson chi-square)	-	<0.001
Gender		
Male	396	64.6%
Female	400	64.0%
p values (Pearson chi-square)	-	0.83
Race/ethnicity		
Hispanic	1282	66.3%
White non-Hispanic	2052	72.8%
Black non-Hispanic	64	63.4%
Asian non-Hispanic	71	65.1%
Other/Multi race, non-Hispanic	51	60.7%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	361	60.0%
Completed high school or higher education	3363	71.1%
p values (Pearson chi-square)		<0.0001
CSHCN status		
Non-CSHCN	3431	69.0%
CSHCN	431	77.0%
p values (Pearson chi-square)	-	<0.001
At risk for developmental delay (online only)		
Low/No risk	1645	73.7%
High/Moderate risk	544	73.4%
p values (Pearson chi-square)	-	0.88

KPNW study: After controlling for other child and family demographic and health factors and provider characteristics, the likelihood (or adjusted odds ratio-AOR) that a child met quality measure criteria differed significantly according to: (1) child's age and (2) child's birth order

Table 2b4.9b: Mean number of developmental services care components for which quality care was received and the proportion of children meeting criteria for receiving quality developmental services by characteristics of children and families.

Characteristic of Child or Child's Family		% Meeting Criteria
Child's Age		
	Less than 9 mos.	44.2% [S]
	10 to 18 mos.	53.8%
		AOR: 1.54
	19 to 49 mos.	56.8%
		AOR: 1.67
Child's Gender		
	Male Child	54.1% [NS]
	Female Child	52.3%
Child's Race		
	White, Non-Hispanic	54.1% [NS]
	Asian, Non-Hispanic	52.7%
	Hispanic	51.3%
	Other Race, Multiple Race	48.6%
Birth Order		
	Not First Born	56.2% [S]
	First Born	50.0%
		AOR: .80
Child's Risk for Developmental, Behavioral or Social Delays (Using Parent's Evaluation of Developmental Status)		
	Low/No Risk	52.9% [NS]
	At Risk	54.7%
Respondent Education		
	More than High School	53.6% [NS]
	High School or Less	50.7%
Respondent's Risk for Depression (Using the Kemper Screener)		
	No Symptoms of Depression	54.4% [NS]
	Symptoms of Depression	49.0%

NOTE: Adjusted odds ratios (AOR) derived from regression analyses listed in the table are shown only if they are statistically significant. AOR uses the first subgroup of each characteristic as a reference.
[S] differences significant at the $p < .05$ level of significance;
[NS] differences not significant.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

The demographic and socioeconomic survey items included in the AAPC quality measure make it possible for providers to identify populations and subgroups for which health service delivery improvement is most needed.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

Not applicable

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Differences in the AAPC scores across providers is demonstrated for (1) 5 top individual providers with the highest number of surveys using Online PHDS data; (2) across 56 providers using KPNW study data; and (3) pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.

Online PHDS: The performance scale for the AAPC quality measure was calculated using the scoring methods described in Attachment A-4. Individual provider level differences in performance were illustrated by the proportion of children meeting the quality of care criteria across 5 top providers with the highest number of completed surveys after their well-child visit.

KPNW Study: The significance of differences observed in the proportion of children meeting criteria for the AAPC quality measure across pediatric providers (n=56) was evaluated using t-tests. The relative spread in the quality measure score across providers was assessed using the coefficient of variation statistics (standard deviation across providers multiplied by 100%). Multi-level regression models were conducted using the pediatric provider as the level 2 clustering variable, in order to assess the degree to which the probability that a child meets criteria on each quality measure is explained by differences between providers (called the "clustering effect"). In implementing this multi-level regression method (Empty Model), the presence of a significant clustering effect by pediatric providers was estimated prior to accounting for the child and family characteristics associated with each provider. Second, variables related to the child and family characteristics (child's age, gender, race/ethnicity, birth order, developmental and behavioral delay risk status; parent education and risk for depression) were added to the Empty Model to assess how much of the provider clustering effect observed remains after accounting for these characteristics (called the "Patient Model").

HRSA study: Quantitative data results for the baseline (2010) and follow-up (2011-12) study of the intervention sites using the HRSA Evaluation Study data were conducted using basic descriptive statistics to describe each sample and applying chi-square test of statistical significance to assess differences in the quality measure for the baseline and follow-up samples.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Online PHDS: Table 2b5.2a present the proportion of children whose care met for the quality measure across 5 providers. The proportion of parents who reported that the health care provider asked parents whether they have

concerns about their child's development and/or behavior ranged 64.9%-76.4%. Only 6%-45% of parents of young children reported that their child's pediatric clinician asked about their concerns regarding their child's development or behavior.

Table 2b5.2a: Proportion of Children Meeting Measure Criteria, Top 5 individual providers with highest number of surveys

Characteristics	All Children (n=5355)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
Asking parents whether they have concerns about their child's development and/or behavior.	69.8%	76.4%	70.3%	69.0%	92.3%	64.9%

KPNW Study: In this study, a little over one-half of children had parents who reported that their child's well-child care provider asked them whether they had any concerns about their child's development and/or behavior (53.3%). The range across providers in the proportion of children who met the AAPC quality measure criteria was substantial and statistically significant (p=0.002).

Table 2b5.2b: Proportion of all children in the study who met criteria for receiving quality services and ranges in proportion across providers. (SD=Standard Deviation)

Developmental Services Quality Measures	Proportion of All Children Meeting Measure Criteria (n = 2173)	Range in the Proportion of Children Meeting Measure Criteria Across 51 Pediatric Providers	Relative Variation (COV) in Measure Scores Across Pediatric Providers
Asking parents whether they have concerns about their child's development and/or behavior.	53.3%	20.0% to 76.0% SD: 11%; (p = 0.002)	20.6%

Only providers with 15 or more completed surveys are included in the provider level analysis. Provider level survey completion numbers range from 15 to 153.

Multi-level analysis: For the Empty Model that used the provider as the level 2 clustering variable, only 1.1% to 2.2% of the total variance observed in whether children met AAPC quality criteria was explained by either measured or unmeasured differences between the providers that they see. This suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers. These findings translate into a 1.19 to 1.29 median odds ratio across all quality measures, including Ask About Parent Concerns, indicating that the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across all quality measures. When child/family level characteristics are added to the model (Patient Model), the total variance explained by differences between providers does not change significantly.

HRSA study

The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the AAPC measure. Parents were more likely to report being asked about their concerns regarding their child's development or behavior at the follow-up assessment than at the baseline assessment; and parents were more likely to be asked about one or more psychosocial (family assessment) topics at follow-up. The tables below present comparison of percent of children whose parents were asked about their concerns and met the AAPC quality care criteria between baseline and follow-up survey data.

Table 2b5.2c. Asking about Parent's Concerns about Development Measure, by Children's Characteristics: Parent was asked if they had concerns about their child's development

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	64.6% (357)	73.7% (235)	0.01
10-18 months	78.6% (319)	76.6% (246)	0.59
19-48 months	80.3% (282)	82.3% (181)	0.59
Race			
Hispanic	72.0% (72)	84.1% (37)	0.14
White	72.5% (745)	76.4% (542)	0.07
Asian	75.0% (21)	70.6% (12)	0.74
Multiple or other	87.5% (21)	77.8% (14)	0.44
Insurance type			
Private or private and public	72.4% (784)	75.5% (509)	0.16
Public only (includes Medicaid, Medicare, CHIP, and Military)	78.4% (149)	82.2% (129)	0.42
Other insurance type	71.4% (5)	(3)	-
Uninsured	50.0% (6)	100% (7)	0.04
At risk of developmental delay			
Low/no risk	72.2% (741)	76.3% (411)	0.09
High/moderate risk	76.2% (214)	82.4% (140)	0.13

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

Significant gaps and unexplained variations remain in the quality of developmental services for young children overall and AAPC in particular. The probability of a parent being asked about their concerns about their child's behavior and development varies nearly as much across children seeing the same provider as across providers. The AAPC quality measure assessed here provides a relatively comprehensive picture of performance in the area of preventive and developmental services for young children.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Table 2: Proportion of all children in the study who met criteria for receiving quality developmental services across six components of care and ranges in proportion across providers and offices. (SD=Standard Deviation)

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The AAPC quality measure items were developed based on several rounds of cognitive interviews with parents to ensure quality of responses appropriate to the questions and minimize missing responses.

Online PHDS: Rate of survey completion was calculated based on survey start and complete dates for each respondent. According to the quality measure scoring protocol, if a parent answered less than half of the items in the quality measure, their score is considered to be missing. This does not include items that should have been appropriately skipped. Missing responses are not given a valid score and are not included in the calculation of the quality measure.

KPNW Study: Of the 5,755 sampled children, 2,173 surveys were returned (37.8%). For these children, the provider the parent identified and the provider to which the child was assigned by the health plan were the same 97.3% of the time. A 95% response rate was obtained for the provider survey.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Online PHDS data show that overall 2.6% of parents who started the survey did not complete the survey. Table 2b7.2a presents the frequency of missing values for the Asking about Parental Concerns measure across all providers and for 5 providers with the highest number of surveys.

Table 2b7.2a. The frequency of missing values for Asking about Parental Concerns, overall and top 5 providers

Quality measures	Overall	Provider ID				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Ask about concerns and addressing concerns	2.4% (138)	4.3 (18)	0	0.3 (1)	0	2.6 (2)

KPNW study: Children whose parents responded were not different from those who did not respond in terms of their gender and insurance type. The responding population were slightly less likely to be in the 19 to 48 month age group (55.8% sampled, 51.5% responding) and were somewhat more likely to have had more than one well-visit in the past (67.5% sampled, 74.7% responding).

Table 2b7.2b. Sociodemographic Characteristics of KPNW Starting and Responding Sample

Characteristic	Proportion of Starting Sample (N=5755)	Proportion Respondents as of (N=2162)
Gender of Child[NS]		
Male child	52.7	53.7
Female child	47.3	46.3
Age of the Child[S]		
Child age 3-9 months	19.4	21.8
Child age 10-18 months	24.9	26.7
Child age 19-48 months	55.8	51.5
Type of Insurance[NS]		
Private	98.6	98.5
Public	1.4	1.5
Child's Health Care Utilization		
Number of well-child visits [S]		
1 Well-Child Visit	32.5	25.3
2 or More Well-Child Visits	67.5	74.7
Number of emergency room/urgent care visits		
0 ER/urgent care visits	49.8	51.0
1 ER/urgent care visit	26.2	25.8
2 or more ER/urgent care visit	24.0	23.2
Number of overnight hospital stays [NS]		
0 overnight hospital stays	96.6	96.9
1 or more overnight hospital stays	3.4	3.1

[S]Denotes variables for which statistically significant variation exists between the starting and responding sample for the target child or respondent characteristic.
[NS]No significant variation exists between the starting and responding sample for the target child or respondent characteristic.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Information about non-respondent is not available to compare with those who responded the survey because online PHDS is publicly available tool. However, the low rate of incomplete survey (2.6%) suggests that the measure was acceptable to respondents. Overall, the quality measure had 2.4% of missing cases, ranging 0-4.3% across the top 5 providers with the highest number of surveys. Few overall missing values suggest that the measure level results unlikely to be biased by non-response to the survey questions.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Other

If other: Data are generated by parents completing the CAHMI-developed Promoting Healthy Development Survey (PHDS), which is sent to them by their provider following a well child visit.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

Patient/family reported information (may be electronic or paper)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

Data are parent-report using the CAHMI developed Promoting Healthy Development Survey (PHDS). CAHMI captures the data at the provider level through a process described above and in the Evidence Form, Figure 1.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.

During 2012-2016 we have experienced some operational delays. In 2012, the provider feedback reports were not automated. When providers wanted a summary report, CAHMI had to manually create them. This was excessively time consuming and CAHMI did not have resources to continue the manual generation of the reports. We sought and received funding to automate the reports. Some difficulties with contractors and staff change-over caused major delays in the project. Then, CAHMI moved from the Oregon Health & Sciences University to Johns Hopkins University School of Public Health in 2014, and it was necessary to upgrade the CAHMI servers. No technical support was available for the transition which caused further delays. Additionally, the PHDS was originally developed in 2001; thus much of the coding and back-end technology for this tool was antiquated and ceased to function after the move. Consequently, and as a result of new improved technology, we have had to redesign the two PHDS related websites - the PHDS toolkit and the parent survey -- as well as the CAHMI PHDS database. Lack of funding caused delays. However, we anticipate launching the new PHDS in February 2017.

Time and cost of data collection are low: provider registration takes about 10 minutes and the parent survey takes about 15-20 minutes to complete. To date, implementation has been limited by lack of funding and resources for outreach, communication and technical support. Our experience in the development and evaluation of the PHDS demonstrated a clear and compelling need to work closely with providers to overcome the many myths that both parents and providers have about patient-engagement quality improvement tools. For the PHDS to be adopted by providers, it is essential to demonstrate, for example, that tool adds value for both the parent and provider, that it fits into and typically improves work flow in the office; improves parent-provider communication, and most important, improve the quality and delivery of nationally recommended services for children. This can only really be accomplished by collaboration and partnership with providers.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Professional Certification or Recognition Program Quality Improvement (external benchmarking to organizations)	

4a.1. **For each CURRENT use, checked above (update for maintenance of endorsement), provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

NA

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
The PHDS toolkit (www.phstoolkit.org) and the parent-reported PHDS (www.wellvisitsurvey.org) were used by 68 uniquely identified providers across the country through 2013. We are happy to provide a list of these providers to NQF if desired. In 2014, CAHMI moved from the Oregon Health & Sciences University, Portland OR to the Johns Hopkins University, Baltimore, MD. As a result of the move, and because both server and database technologies had rapidly evolved and improved over the past few years, it was necessary to upgrade our servers, which in turn caused some technical issues with the links between the provider toolkit, the PHDS, and the CAHMI PHDS database. Additionally, the PHDS was originally used to compare providers within a practice as well as between practices within a health system. The anticipated use of the Online PHDS is intended to provide feedback only for individual providers but not between providers. The combination of these factors led to a decision to upgrade and redesign the PHDS toolkit, PHDS database and Parent Survey. (The PHDS parent survey itself, however, remains fully operational, although use has been nominal from 2014-present, and can be accessed at www.wellvisitsurvey.org.) The redesign

required additional time, IT and CAHMI staff resources and delays were incurred during 2014-2015. However, we are now in the process of finalizing the PHDS Toolkit and database redesign, which is anticipated to be completed and launched in February 2017.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
The redesigned PHDS system (registration toolkit, parent survey tool and PHDS database) is anticipated to be completed and fully functional by February 2017. We have a communication and outreach plan to promote the PHDS as part of the CAHMI Cycle of Engagement, see Attachment A-5), which includes the CAHMI Well Visit Planner (www.wellvisitplanner.org) -- a free parent engagement tool that helps prepare parents for the upcoming well child visit – and the post-visit PHDS which assesses whether the parent received services in alignment with national guidelines as well as family centered care. We have been promoting the Cycle of Engagement in national meetings (AMCHP, PAS, APHA, AcademyHealth ARM, National Child Heath Policy Meeting, and more) over the past several years. We presented the Cycle of Engagement at the CMS Quality Meeting December 13, 2016 and have further plans to unveil the redesigned version at meetings in 2017. The WVP and PHDS have also been endorsed tools that meet requirements for Bright Futures implementation.

We have received substantial interest in the CAHMI parent-engagement tools (both the WVP and the PHDS) from and are in extensive conversations with a number of organizations and agencies including health systems, payers, provider organizations – (CMS/Medicaid, Title V, Head Start, Kaiser Permanente and others); professional associations such as the American Academy of Pediatrics, Bright Futures, National Medicaid Medical Directors, the Children's Hospital Association (CHA), AcademyHealth, Association of Maternal and Child Health Programs (AMCHP), CityMatCH, National Initiative for Children's Healthcare Quality (NICHQ), Autism Speaks, Prevent Child Abuse America; National Prevention Information Network (NIPN); national community-based programs and organizations; philanthropic funders; software platform and electronic medical records systems developers and family organizations. We are in the process of securing funding for Cycle of Engagement EMR integration and implementation projects in partnership with or from a number of interested parties. Further, we are finalizing our application to the American Board of Pediatrics to have the Online PHDS certified as a web-based Maintenance of Certification (MOC) (Part 4) quality improvement (QI) tool for pediatricians. ABP has expressed significant interest in the PHDS and provided some initial funding for the redesign efforts.

Improvement
Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
Based on PHDS feedback results from an evaluation of the WVP conducted in 2011-2012 in Oregon, we found that Asking about parental concerns did not change significantly because the quality on this measure was already high: Baseline assessment (2010) for this measure was 84.3% and post-assessment testing showed about the same, 83.3% (AOR:0.94, CI: 0-58-1.54).

4c. Unintended Consequences
The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**
There were no unintended or unexpected consequences that we are aware of.

4c.2. **Please explain any unexpected benefits from implementation of this measure.**
There were no unexpected benefits that we are aware of.

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
Extensive qualitative interviews with providers and parents have been conducted and previously reported (See Attachment 2, Evidence Report)

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
Key informant interviews and focus groups with parents and providers were held throughout the testing and evaluation period. We obtained baseline and post-implementation information from providers and post-implementation information from parents. It was necessary to work closely with practices to demonstrate value of the family engagement tools (Well Visit Planner and PHDS) as well as to modify the process to fit individual practice office culture and work flow. A significant amount of provider and staff education was needed to overcome fears and myths that the tool would add to, not help, time management and that parents would not want to participate. This was accomplished by continued and persistent relationship building, spending much time in the office setting with the staff and providers and holding frequent Q&A sessions as the process unfolded.

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.
Through key informant interviews and focus groups with parents and providers.

4d2.2. **Summarize the feedback obtained from those being measured.**
The Ask and Address Parental Concerns measure is seen by providers as an excellent way by which they can improve the quality of the well child visit. In particular this matters a great deal to the providers who are being financially incentivized for family-centered care outcomes.

4d2.3. **Summarize the feedback obtained from other users**
For the most part, parents appreciated being asked about their experience with their well child visits and used it as a way to provide confidential feedback to the providers.

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
The feedback was helpful for future implementation efforts of CAHMI's family engagement tools. The feedback, however, did not result in any changes to the measure itself.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization of Related Measures
The measure specifications are harmonized with related measures;

OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications harmonized to the extent possible?
Yes

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
NA

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
NQF #0011 - the PHDS (Promoting Healthy Development Survey) - was endorsed by NQF on October 4, 2012. The PHDS contains the Ask About Parental Concerns measure. Neither the questions nor the scoring of the questions have changed since the PHDS was endorsed. It is not actually a competing measure; rather, the Ask About Parental Concerns measure is embedded in the PHDS tool.

Please note: The PHDS endorsement (#0011) can be found on the NQF measures website but does not appear to be found in the NQF directory in Question 5 above. Hence, we were forced to enter a "no" to Q5 in order to submit this application.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Attachment_A_Supplemental_Materials_Revised_01_18_17-636203525678362747.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Child and Adolescent Health Measurement Initiative
Co.2 **Point of Contact:** Christina, Bethell, cbethell@cahmi.edu, 443-287-5092-
Co.3 **Measure Developer if different from Measure Steward:** Child and Adolescent Health Measurement Initiative
Co.4 **Point of Contact:** Christina, Bethell, cbethell@cahmi.edu, 443-287-5092-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
National Advisors for Patient Centered Quality Improvement of Well-Child Care:
Betsy Anderson, Family Voices
David Bergman, Stanford University
Dimitri Christakis, University of Washington
Paula Duncan, University of Vermont
Cynthia Minkovitz, Johns Hopkins School of Public Health
Amy Perritti, American Academy of Pediatrics
Ed Schor, The Commonwealth Fund

Judy Shaw, University of Vermont
Sara Slovin, Johns Hopkins Medicine

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2002

Ad.3 **Month and Year of most recent revision:** 01, 2017

Ad.4 **What is your frequency for review/update of this measure?** 3 years

Ad.5 **When is the next scheduled review/update for this measure?** 01, 2018

Ad.6 **Copyright statement:** None

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3221
Corresponding Measures:
Measure Title: Family Centered Care
Measure Steward: Child and Adolescent Health Measurement Initiative
Brief Description of Measure: This measure is used to assess the average percentage of recommended of aspects of family-centered care (FCC) regularly received by the parent from the pediatric clinician. Topics specifically focus on the following components of FCC:
(1) whether the health care provider understands specific needs of child and concerns of parent;
(2) builds confidence in the parent;
(3) explains things in a way that the parent can understand; and
(4) shows respect for a family´s values, customs, and how they prefer to raise their child.
Developer Rationale: Family-centered care (FCC) is an integral part of the preventive and developmental services recommended by the American Academy of Pediatrics (AAP), as well as an element of medical home. This measure, as part of the Promoting Health Development Survey (PHDS), captures parent-reported information about the communication and partnership between the provider and the parent that compose FCC which could not otherwise be obtained through medical records or administrative data. Few quality measures have been available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child´s health care. The PHDS provides direct feedback from parents about the delivery and quality of preventive services for their children. The PHDS was developed for the purpose of assisting providers, consumers, purchasers, and policymakers in assessing the degree to which health plans and practitioners provide developmental services as recommended in guidelines set forth by the AAP and the Maternal and Child Health Bureau´s Bright Futures initiative.

Numerator Statement: The numerator measures the number of parents who had a well child visit within the last 12 months and who experienced family centered care in 7 specific areas.
Denominator Statement: The denominator is the number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months who answered the Family Centered Care questions on the Promoting Healthy Development Survey (see Attachment A-2, page 12).
Denominator Exclusions: Missing data for the Family Centered Care questions excluded from analysis.

Measure Type: Outcome: PRO
Data Source: Other
Level of Analysis: Clinician : Individual

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report
1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary

- This is a Patient-Reported Outcome-Based Performance Measure (PRO-PM) derived from the responses to seven questions on the Promoting Healthy Development Survey (complete survey starts on page 20 of the Appendix).
- The developer provided a logic model in both graphic and narrative: (1) the parent and child attend a well child visit with their provider; (2) the provider subsequently sends a survey -- the Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org), which includes one question (7 items) on Family Centered Care (see Attachment A-2, page 12) for the parent to complete; (3) when at least ten surveys have been completed, the provider receives a feedback report on parents' experiences of the visit and the extent to which they felt they received family centered care via the CAHMI PHDS Toolkit website (www.phdstoolkit.org); (4) the provider reviews the report and then can engage in a Plan-Do-Study Act (PDSA) quality improvement process to improve their Family Centered Care quality score.
- The developer also provides the following evidence:
    - Family-centered care (FCC) is an integral part of the preventive and developmental services recommended by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau as part of the Bright Futures guidelines, as well as an element of medical home.
    - The process outlined in the logic model allows health care providers to better understand the extent to which their patients experience "quality care" – in this case, the extent to which parents felt they received family centered care. It also allows providers to engage in quality improvement activities to improve their parent-reported Family Centered Care quality score by using several Plan-Do-Study Act (PDSA) cycles.
- The developer notes in the Performance Gap section a HRSA study "found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to report their needs met for family centered care at the follow-up assessment than at the baseline assessment; and parents were more likely to be asked about one or more psychosocial (family assessment) topics at follow-up." The results are included in the testing attachment.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Guidance from the Evidence Algorithm: Patient-reported outcome (Box 1)→ Relationship between PRO and provider action (Box 2)→ Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer reports the top 5 individual providers' performance scores ranged from 78.9%-89.3%.
- A Kaiser Permanente Northwest (KPNW) Study found that 83.9% parents reported that their child's well child care met family-centered care criteria. The variation across providers was statistically significant (p=<0.001).
- A HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures.

Disparities

- The online PHDS found that variation is observed according to a child's age; race/ethnicity (Hispanic=78.7%, white=82.9%, black=74.7%, Asian=66.0%, other/multirace=76.9%, p<0.0001); level of risk for developmental, behavioral, or social delays (low/no risk=84.1%, high/moderate risk=74.3%, p<0.0001) across all quality measures. Non-Hispanic white children are more likely to meet criteria on the Family Centered Care measure. Children of lower educated mothers are less likely than those with more educated mothers to have high Family Centered Care. Non-CSHCN and children high/medium risk are more likely to receive care met family centered care criteria.
- The KPNW study found that a family receiving high quality family centered care differed significantly according to: (1) child's race/ethnicity for four (white=86.6%, Asian=64.0%, Hispanic=80.3%, other/multiple race=83.1%) (2) child's birth order (not first born=86.7%, first born=81.6%), (3) child's developmental and behavioral risk status (low/no risk=86.3%, at risk=78.6%), (4) respondent education level (more than high school=84.9%, high school or less=78.1%) and (5) parent risk for depression (depression symptoms=77.8%, no symptoms=85.6%).

Questions for the Committee:

o *Is there a gap in care that warrants a national performance measure?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

- It was reassuring to see the references to AAP, MCH, Bright Futures, and HRSA. As this is an outcome measure, perhaps relating FCC to improved health outcomes would be preferable. As the numerator refers to 7 areas, the brief description measure should mention this (only covers 4.) The ages of the children need to be consistent in the numerator and denominator and if so then agree that the measure passes.
- Evidence was provided that by measuring the answers to the questions and then doing "improvement activities" the practitioner can raise their score on these items.
- It is originally unclear how the top 5 providers where selected (explained later in the document). There needs to be clarification on why "lower educated mothers" are less likely to "have high family centered care". This should be related to the provider, not parent (and there is no consideration of fathers or other guardians such as grandparents), for consistency. Rating: moderate.
- The HRSA study that was cited does not include measures of sustainability long after the Bright Futures Guidelines training session. Was the improvement after the training a one time event, or has this improvement been sustained? The developer did not address the under representation of black/African American population (in fact, only one of the entities included any numbers for this population); other groups that seem under-represented in the use of the questions by KP, HRSA, and PHDS are high school or less than high-school education and publically insured (Medicaid/CHIP). Also, the survey was limited to those who can read and understand English- thereby eliminating a population that would benefit from patient centered care.
- There are performance gaps in care as identified by these measures. The gaps are based on race/ethnicity, parental educational level, child's risk for developmental problems and parent risk for depression.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

- Level of analysis: Clinician – individual
- Interpretation of score: Better quality = Higher score
- This is a patient-reported outcome-based performance measure (PRO-PM)
- Numerator: *The numerator measures the number of parents who had a well child visit within the last 12 months and who experienced family centered care in 7 specific areas.*

- Denominator: *The denominator is the number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months who answered the Family Centered Care questions on the Promoting Healthy Development Survey (see Attachment A-2, page 12).*
- Exclusions: The developer states that "Missing data for the Family Centered Care questions excluded from analysis." [NQF does not consider this an exclusion as it is defining the population of the measure.]
- The developer includes a calculation algorithm.
- The measure is not risk adjusted or risk stratified, but the developer states that it can be stratified by variables such as child demographics characteristics (e.g., the child´s age, race); child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or parent health characteristics, if large enough data sets are available.
- The measure does not use sampling.
- This measure relies on a set of questions within the Promoting Healthy Development Survey. This online survey is initiatied by the provider who sends it to a parent after a well-child visit. Providers must have a minimum of 10 surveys to generate a report to maintain parent confidentiality.

Questions for the Committee:

- o *Are all the data elements (question items) clearly defined?*
- o *Is the logic or calculation algorithm clear?*
- o *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

NQF Note: Both measure score and data element reliability testing are required for PRO-PMs.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

Method(s) of reliability testing

- The developer used data from the online Promoting Healthy Development Survey (PHDS), a Kaiser Permanente Northwest Study (KPNW), and a HRSA evaluation study that tested "three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young children in the context of discussions between pediatric clinicians and parents during well-child visits"; the HRSA study used 5 tools including the PHDS.
- The Chronbach alpha to test internal consistency (data/item element reliability) was calculated using the Online PHDS and KPNW data. In addition, factor analysis was performed to investigate the dimensionality of the scale.
- To test the survey itself, inter-item correlation was assessed to insure redundancy of the questions.
- Score level reliability was assessed by inter-unit reliability (IUR) using analysis of variance.

Results of reliability testing

- Using the top 5 individual providers with the highest number of surveys (N=77 to 94) from the online PHDS testing, the developer reports the Chronbach's alpha for internal consistency (item-level) range from 0.81-0.95 with the mean score for all providers at 0.88. In the KPNW study, the top 5 individual providers (n=66 to 80) had Chronbach's alphas of 0.77-0.90, with one having too small a sample to provide results; the mean score for all providers was 0.81.
- The developer reports that the results for the inter-unit reliability (IUR) testing are within the recommended threshold (0.73) to reliably demonstrate differences between providers.
    - o [*NQF note: IUR measures the proportion of the measure variability that is attributable to the between facility variance. A small IUR (near 0) reveals that most of the variation of the measures between*

facilities is driven by random noise, indicating the measure would not be a good characterization of the differences among facilities, whereas a large IUR (near 1) indicates that most of the variation between facilities is due to the real difference between facilities. The recommended range is above 0.70.]

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm: Precise specifications (Box 1)→ Empirical reliability testing (Box 2) → Score level testing (Box 4) → Appropriate method used (Box 5) → High certainty or confidence that the performance measure scores are reliable (Box 6a)→ High.

Highest possible rating is HIGH.

Note: PRO-PMS *require* element-level testing as well, which was conducted. If judged without score-level testing, the highest eligible rating for this type of testing is MODERATE.

Preliminary rating for reliability: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

NQF Note: Both measure score and data element validity testing are required for PRO-PMs.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity**
- ☒ **Empirical validity testing of the measure score**

Validity testing methods:

The developer conducted several methods of validity testing:

- Factor analysis was conducted to assess the construct validity of the quality measure.
- To assess the concurrent validity of the measure scare, the developer tested the hypothesis: "Respondents who indicate that providers talked with them about recommended anticipatory guidance topics or providers who discussed family psychosocial issues are more likely to report receiving family-centered care."
- Pearson correlation coefficients were calculated across all scale measures to test hypotheses about expected relationships among the PHDS quality measures and to assess the degree to which each of the PHDS quality

measures provide unique information. The developer notes that "We expect a moderate or strong correlation between the family assessment scale measures (>0.30) and inter-scale correlation coefficients to be less than 0.80."
- The PHDS survey also was tested using focus groups, in-depth cognitive interviews, a literature review, and an advisory board of expert stakeholders.

Validity testing results:
The developer reports the following results:
- "Average factor loading for FCC was 0.70. Inter-item correlation ranged between 0.55-0.68. Factor analysis suggests that the scale items are unidimensional." Acceptable ranges for factor loading are above 0.60.
- The concurrent validity testing results demonstrated that parents were more likely to receive family centered care if they also reported that their questions on specific anticipatory guidance topics were answered or if the provider discussed family psychosocial issues.
- The developer provides a table of Pearson Correlation Coefficients, which assesses whether the measures are examining different topics. The results suggest, according to the developer, that the measures are not redundant, with an average correlation of 0.34. This measure was most highly associated with the *Anticipatory guidance and parent education measure* (0.52).

Questions for the Committee:
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:
N/A

Questions for the Committee:
- *Is the lack of exclusions consistent with the evidence?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included ? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary
- The developer does not risk adjust the measure because "we do not expect variation in the quality of family centered care provided for children due to risk factors, e.g. children with special health care needs. The performance should be the same regardless of risk factors."
- The developer notes the measure can be stratified by several demographic or health variables as "Identification of variation in quality measures across subgroups of children helps to highlight aspects of care and population of children for which preventive and developmental services may be most need of improvement."
- The developer reports that many studies have shown differences in access to and quality of care, as well as parent satisfaction. The developer states that "One study found: Overall, 94.0% of parents reported 1 or more unmet needs for a number of aspects of care, including assessing family alcohol use, substance abuse and safety. Uninsured children and children aged 18 to 35 months are disproportionately represented among the 15.3% of children whose parents indicated an unmet need this area of care. There are significant variations in performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors."
- Variations were observed by demographic and socioeconomic factors.

Questions for the Committee:
- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- The developer was able to demonstrate meaningful differences among providers for the top 5 providers (number of individual surveys completed) in the online PHDS; across 56 providers using KPNW study data; and pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.
- Online PHDS: For the top 5 individual providers with the highest numbers of surveys (n=77 to 91), a range of 78.9%-89.3% of parents of young children reported receiving family centered care; the average for all children was 79.8%.
- KPNW Study: 83.9% of children had parents reporting that they received family centered care. Range across providers in the proportion of children who met quality measure criteria was statistically significant, 52.8%-95.8%. Provider n ranged from 15-153.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

The developer reports the following:

- **Online PHDS:** Rate of survey completion was calculated based on survey start and complete dates for each respondent. According to the quality measure scoring protocol, if a parent answered less than half of the items in the FCC measure, his/her score is considered to be missing. This does not include items that should have been appropriately skipped. Missing responses are not given a valid score and are not included in the calculation of the quality measure.
- **Online PHDS** data show that 2.6% of parents who started the survey did not complete the PHD survey. For the family centered care questions, the frequency of missing values averages 3.8% for all providers, and for the top 5 providers (n=75 to 91) the range was 0%-3.2%.
- **KPNW Study:** Of the 5,755 sampled children, 2,173 surveys were returned (37.8%). For these children, the provider the parent identified and the provider to which the child was assigned by the health plan were the same 97.3% of the time. A 95% response rate was obtained for the provider survey.
- The developer notes that responses for the KPNW survey did not differ by gender or insurance type, but did differ by age and by number of previous well visits.
- The specifications indicate that "Surveys missing four or more of the responses to the Family Centered Care questions are excluded from analysis." However, no information is provided on why this does not bias the responses; the online PHDS indicates an average of 3.8% of surveys had missing data for this section.
- Information about non-respondents is not available, but "Overall, the quality measure had less than 4% of missing cases, ranging 0-3.2% across the top 5 provider with highest number of surveys. Few overall missing values suggest that the measure level results unlikely to be biased by non-response to the survey questions."

Guidance from the Validity Algorithm: Specifications consistent with evidence (Box 1)→ Threats to validity addressed (Box 2) →Empirical validity testing (Box 3)→ Measure score testing (Box 6)→ Appropriate method (Box 7)→ Moderate certainty or confidence that the performance measure scores are a valid indicator of quality (Box 8b)→ Moderate

The highest possible score is MODERATE.

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
RATIONALE: Missing data is not adequately addressed; non-respondent bias not available.

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b)

- Patient-reported data is not as accurate for outcome measures. The clarification that NQF doesn't consider the missing data to be an exclusion is appreciated. It is not explained why this measure does not use sampling (though it is explained for similar measure 3220). It is agreed that the data elements are clearly defined, the calculation algorithm is clear, and this measure can be consistently implemented.
- The calculation algorithm uses only the answer "yes, and my questions were answered, however, the answer "no, but I already had information on that topic and didn't need to talk about it" would seem to be family centered as well since it would be consistent with providing families with the information that they need/want.
- It was reassuring that the reliability testing measure including both measure score and data element. It was also reassuring that the Reliability testing performed with the data source and level of analysis indicated. The results of the reliability testing explain how the top 5 providers where selected but this should have also been explained in the previous section as mentioned above for clarity. There was concern with one sample size being too small. Rating: moderate.
- Please address concerns about using data from a 2004-2005 study (more than 10 years old). Also, without a control group, how do you know the improvement would not have occurred during the timeframe of each study. When were the focus groups conducted and what was the make-up of the participants?
- Ten surveys per practitioner is not very many to give a good picture of the average performance of a practitioner. There are a number of entities that have used these questions in some format although not clearly all in the same manner as presented here (e.g. Massachusetts only used some of the questions). That one of the top 5 providers who had a range of 77 to 96 surveys each didn't have enough to provide results is a bit of an issue since it is unclear what the other providers' numbers looked like and how many providers had full results since the others had less than these did.
- There is no section to reply for 2b1 "validity specifications"-it is agreed that the specifications are consistent with the evidence. For section 2b2, it was reassuring that the validity testing level used both measure score and data element testing against a gold standard. It was further reassuring that empirical validity testing was done in addition to face validity. It is noted that the inter-item validity testing started in the low range of .55 which should be .60 minimum. Again, as an indicator of quality FCC should be linked with improved health outcomes.
- The validity appears to be adequate although the numbers of surveys could impact that. This measures quality as defined by family care. The measures don't appear to be duplicative and therefore are measuring different aspects of provision of quality care in the realm of family centered care.
- For 2b3, the lack of exclusions is consistent with the evidence as previously NQF determined the missing data didn't count as an exclusion. The acronym SDS should be defined under 2b4. There seems to be a contradictory statement to say "performance should be same regardless of risk factors " then state "there are significant variations in performance". For 2b5, it is agreed that the data demonstrates meaningful differences in quality. For 2b6, it is unclear why comparability isn't applicable. For 2b7, it is understood that NQF didn't consider the missing data as an exception however this is a concern regarding missing responses and bias. Rating: moderate.
- Not clear if these questions are still valid after so many years after being tested. Under representation of payer mix and racial/ethnic diversity a concern.
- Evidence shows that children and families with increased risk factors including social determinants are less likely to receive quality care and more likely to have unmet needs. However, the expectation for provision of quality care should be the same for all children and families. If a practitioner provides the same level of care to all patients, then there should be no need for risk adjustment. There is a clear range of performance between providers indicating that the results of the surveys can detect a meaningful difference in performance. Non-respondents could be an issue if one can't identify whether or not there are differences between families that responded and those that didn't. Surveys with missing data could be analyzed by question using the information that was provided.

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- This is a patient reported outcome. Data are generated by parents completing the CAHMI-developed Promoting Healthy Development Survey (PHDS), which is sent to them by their provider following a well child visit.
- Although the survey has been in use since 2001, there is not currently an automated reporting system for providers. The developer has been working on a new website for the survey that will automatically report data, and expects it to launch in February 2017.
- The developer reports that the provider registration takes about 10 minutes and the parent survey takes about 15-20 minutes. There are no fees, licensing requirements, etc, to use the measure.

Questions for the Committee:

- *Is the data collection strategy ready to be put into operational use?*
- *Does the developer have a status update on the new website?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

- Again, patient reported data are less accurate for outcome measures. This is no automated reporting system. It is unclear if the website met the 2/17 deadline. If so, then rating would be moderate; if not then rating would be low.
- Patients are in-undated with satisfaction/patient experience surveys. Is the developer certain there would be a reasonable response-rate if administered now?
- This has been used for a number of years, but it is a bit of a burden for families (15 to 20 minutes to complete the survey). An electronic system could be used to not only gather the data, but also to analyze it. However, it likely won't change the family impact in terms of answering the questions.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No** ☐ **UNCLEAR**

OR

Planned use in an accountability program? ☐ **Yes** ☒ **No**

Accountability program details

No confirmed use for an accountability program, but the developer has been in discussion with a number of organizations that are interested in using the measure, including CMS/Medicaid, Title V, and Head Start.

Improvement results

The developer reports limited improvement over the study period: "Based on PHDS feedback results from an evaluation study conducted in 2011-2012 in Oregon, family centered care did not significantly improve between baseline assessment (69.4%, n=370) and post-intervention assessment (70.1%, n=176, AOR: 1.08, CI: 0.75-1.57), in part because levels of family centered care in this population was already relatively high."

Unexpected findings (positive or negative) during implementation

The developer was not aware of any unintended consequences.

Potential harms

The developer was unaware of any potential harms.

Vetting of the measure
N/A

Feedback:

N/A

***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?*
- *How has the measure been vetted in real-world settings by those being measure or others?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 4: Usability and Use

- It is a concern that this measure is not publicly reported or part of an accountability system. Although FCC initially was high, it is concerning that there was no statistically significant improvement between baseline and post-intervention assessment. It is unclear why vetting is listed as not applicable. Regarding furthering the goal of high quality health care, again linking FCC with improved outcomes would be preferable. There is no space for section 5 so will respond here. It is agreed that there are competing measures, most notably measure 3220. However, as this measure was not vetted, it is not eligible for endorsement and designation.
- The measure has been used in some insurance plans and Medicaid programs. Theoretically practitioners can use this information to improve how they deliver family centered care. However, some of the users of this survey did not show improvement in care despite having the results of this. The developers assumed that this was because there was already high levels of quality care. Unfortunately, the % of family centered care provided was only 70% which would seem to be able to be improved.

Criterion 5: Related and Competing Measures

Related or competing measures

This measure is part of a set of five based on the PHD survey.

- 3219: Anticipatory Guidance and Parental Education
- 3220: Ask About Parental Concerns
- 3221: Family Centered Care
- 3222: Assessment of Family Alcohol Use, Substance Abuse and Safety
- 3223: Assessment of Family Psychosocial Screening

Harmonization

N/A

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as

demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

RATIONALE IF NOT ELIGIBLE: The measure has not been vetted by those being measured or other users.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2965
Measure Title: Family Centered Care Click here to enter measure title
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: 1/13/2017

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome
☒ Health outcome: Click here to name the health outcome
☒Patient-reported outcome (PRO): Family Centered Care

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Figure 1 (attached) shows the logic model by which the Family Centered Care quality measure is obtained and improved. Simply said: (1) the parent and child attend a well child visit with their provider; (2) the provider subsequently sends a survey -- the Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org), which includes one question (7 items) on Family Centered Care (see Attachment A-2, page 12) for the parent to complete; (3) when at least ten surveys have been completed, the provider receives a feedback report on parents' experiences of the visit and the extent to which they felt they received family centered care via the CAHMI PHDS Toolkit website (www.phdstoolkit.org); (4) the provider reviews the report and then can engage in a *Plan-Do-Study Act* (PDSA) quality improvement process to improve their Family Centered Care quality score. THE PDSA cycle involves reviewing the baseline data; developing and implementing a plan of action to improve the score; obtaining further data from the parent; and comparing the first set of results with the second. The full process is repeated until providers are satisfied with their improved scores. We are currently applying for this process to be approved by the American Board of Pediatrics (ABP) for maintenance of certification (MOC, Part 4) credit. The provider must complete three PDSA cycles. Each time point must have at least 25 completed surveys and there must be at least 8 weeks between time periods.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

Family-centered care (FCC) is an integral part of the preventive and developmental services recommended by the American Academy of Pediatrics (AAP), as well as an element of medical home.[1] This measure, as part of the Promoting Health Development Survey (PHDS), captures parent-reported information about the communication and partnership between the provider and the parent that compose FCC which could not otherwise be obtained through medical records or administrative data. Few quality measures have been available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child's health care. The PHDS provides direct feedback from parents about the delivery and quality of preventive services for their children. The PHDS was developed for the purpose of assisting providers, consumers, purchasers, and policymakers in assessing the degree to which health plans and practitioners provide developmental services as recommended in guidelines set forth by the AAP and the Maternal and Child Health Bureau's Bright Futures initiative. The process outlined in the logic model (1a.12) allows health care providers to better understand the extent to which their patients experience "quality care" – in this case, the extent to which parents felt they received family centered care. It also allows providers to engage in quality improvement activities to improve their parent-reported Family Centered Care quality score by using several Plan-Do-Study Act (PDSA) cycles, as described above.

[1] *Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents*, 3rd Edition,

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

Figure_1_Family_Centered_Care_Logic_Model.docx,CAHMI_Family_Centered_Care_evidence_attachment_revised_02_02__17_Final.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

No

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*
IF a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)
IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.

Family-centered care (FCC) is an integral part of the preventive and developmental services recommended by the American Academy of Pediatrics (AAP), as well as an element of medical home. This measure, as part of the Promoting Health Development Survey (PHDS), captures parent-reported information about the communication and partnership between the provider and the parent that compose FCC which could not otherwise be obtained through medical records or administrative data. Few quality measures have been available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child´s health care. The PHDS provides direct feedback from parents about the delivery and quality of preventive services for their children. The PHDS was developed for the purpose of assisting providers, consumers, purchasers, and policymakers in assessing the degree to which health plans and practitioners provide developmental services as recommended in guidelines set forth by the AAP and the Maternal and Child Health Bureau´s Bright Futures initiative.

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis**. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

see also Testing Form-Family Centered Care

DATA SOURCES

Differences in the quality measure scores across providers is demonstrated for (1) 5 top individual providers with the highest number of surveys using Online PHDS data; (2) across 56 providers using Kaiser Permanente NW study data; and (3) pre-post changes across time (2010-2012) after small invervention using HRSA study data for illustrative purpose.

Online PHDS: The performance scale for the FCC quality measure was calculated using the scoring methods described in the Attachment . Individual provider level differences in performance were illustrated by the proportion of children meeting the quality of care criteria across 5 top providers with the highest number of completed surveys after their well-child visit.

KPNW Study: The significance of differences observed in the proportion of children meeting criteria for the FCC quality measure across pediatric providers (n=56) was evaluated using t-tests. The relative spread in the quality measure score across providers was assessed using the coefficient of variation statistics (standard deviation across providers multiplied by 100%). Multi-level

regression models were conducted using the pediatric provider as the level 2 clustering variable, in order to assess the degree to which the probability that a child meets criteria on each quality measure is explained by differences between providers (called the "clustering effect"). In implementing this multi-level regression method (Empty Model), the presence of a significant clustering effect by pediatric providers was estimated prior to accounting for the child and family characteristics associated with each provider. Second, variables related to the child and family characteristics (child's age, gender, race/ethnicity, birth order, developmental and behavioral delay risk status; parent education and risk for depression) were added to the Empty Model to assess how much of the provider clustering effect observed remains after accounting for these characteristics (called the "Patient Model").

HRSA study: Quantitative data results for the baseline (2010) and follow-up (2011-12) study of the intervention sites using the HRSA Evaluation Study data were conducted using basic descriptive statistics to describe each sample and applying chi-square test of statistical significance to assess differences in the quality measure for the baseline and follow-up samples.

PERFORMANCE RESULTS

Online PHDS: Table 4a present the proportion of children whose care met for the quality measure across 5 providers. The proportion of parents who reported receiving family-centered care ranged 78.9%-89.3%.

Table 1B.2a: Proportion of Children Meeting Measure Criteria, Top 5 individual providers with highest number of surveys

Characteristics (n=5355)	All Children					
	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)					
	948 (n=91)	1067 (n=90)	1030 (n=77)	802 (n=75)	1022 (n=75)	
Received family centered care	79.8%	85.7%	80.9%	83.1%	78.9%	89.3%

KPNW Study: 83.9% parents reported that their child's well child care met family-centered care criteria. The variation across providers was statistically significant (p=<0.001).

Table 1B.2b: Proportion of all children in the study who met criteria for receiving quality services and ranges in proportion across providers. (SD=Standard Deviation)

Developmental Services Quality Measures (n = 2173)	Proportion of All Children Meeting Measure Criteria	Range in the Proportion of Children Meeting Measure Criteria Across 51 Pediatric Providers	Relative Variation (COV) in Measure Scores Across Pediatric Providers
Received family centered care	83.9%	52.8% to 95.8% SD: 9% (p = <0.001)	10.7%

Only providers with n=15 or more PHDS responses are included in the provider level analysis. Provider level n ranges from 15 to 153.

HRSA study

The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to report their needs met for family centered care at the follow-up assessment than at the baseline assessment; and parents were more likely to be asked about one or more psychosocial (family assessment) topics at follow-up. The tables below present comparison of percent of children who received care met the quality care criteria between baseline and follow-up survey data for each measure and overall composite comprehensive care measure.

Table 1B.2c. Family Centered Care Measure Comparison by Children's Characteristics

Parent received family-centered care Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	61.9% (343)	65.1% (209)	0.38
10-18 months	64.9% (261)	69.3% (224)	0.23

19-48 months	65.2% (227)	68.0% (151)	0.53	
Race				
Hispanic	62.0% (62)	71.1% (32)	0.35	
White	64.7% (667)	67.7% (485)	0.20	
Asian	39.3% (11)	70.6% (12)	0.07	
Multiple or other	66.7% (16)	66.7% (12)	1.00	
Insurance type				
Private or private and public	64.9% (704)	67.9% (461)	0.20	
Public only (includes Medicaid, Medicare, CHIP, and Military)	58.1% (111)	66.9% (107)	0.10	
Other insurance type	(4)	(2)	N/A	
Uninsured	66.7% (8)	(4)	N/A	
At risk of developmental delay				
Low/no risk	65.5% (669)	69.4% (379)	0.12	
High/moderate risk	57.4% (159)	57.9% (99)	0.92	

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

NA

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

See also Testing Form-Family Centered Care

DATA SOURCES

We used the following data sources for testing of the FCC quality measure:

(1) Online Promoting Healthy Development Survey (PHDS) – data collected through an online, publicly available tool (Promoting Healthy Development Survey-PHDS). Parents who had a well-child care visit in the last 12 months can complete the PHDS. Providers initiate the survey. (See Evidence Form Figure 1 for visual model of the Online PHDS.)

2) Kaiser Permanente Northwest (KPNW) Study – CAHMI partnered with Kaiser Permanente Northwest in Portland, Oregon. The study aimed to evaluate the level and variations in the quality of preventive and developmental services for young children and assess the contribution of key system, provider and patient factors.

STUDY POPULATION

Online PHDS: Children age 3-48 months of age whose parents completed the online publicly available PHDS were included in the testing. During 2008-2016, we received 5,670 completed surveys. Of those surveys, 5,355 surveys with visited provider IDs were used for analyses. Children's socio-demographic and health characteristics varied across the individual providers included in the analysis.

Table 1b.4a: Characteristics of children for whose visited provider ID is available

Characteristics (n=5355)	All Children	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
	948 (n=91)	1067 (n=90)	1030 (n=77)	802 (n=75)	1022 (n=75)	
Age of child						
Under 10 months of age	38.3%	49.5%	33.3%	24.7%	49.3%	34.7%
10 to 18 months of age	34.7%	38.5%	38.9%	57.1%	34.7%	36.0%
19-47 months of age	27.0%	12.1%	27.8%	18.2%	16.0%	29.3%
Race/ethnicity of child						
White, non-Hispanic	53.8%	81.%	20.3%	17.3%	94.4%	53.6%
Hispanic	40.8%	14.3%	74.7%	78.7%	2.8%	37.7
Other race/ethnicity	5.3%	4.8%	7.0%	4.0%	2.8%	8.6%

Respondent education level												
Did not complete high school	12.1%	0	34.1%	15.8%	1.4%	8.1%						
Completed high school	88.9%	100%	65.9%	84.2%	98.6%	91.9%						
Children who qualify for Children with Special Health Care Needs (CSHCN) Screener criteria												
CSHCN	10.1%	8.8%	10.0%	5.2%	21.3%	14.7%						
Non-CSHCN	89.9%	91.2%	90.0%	94.8%	78.7%	85.3%						
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	22.7%	24.4%	-	0%	25.3%	-						

-Data are not available due to small sample size

KPNW Study: The population studied was children 3 to 48 months old who live in a metropolitan area in the Pacific Northwest. One randomly selected child per household whose age would be no younger than 3 months of age and no older than 48 months of age at the time that their parents received the survey and had one or more well-child visits were eligible to be sampled. A random sample of 5,755 children were identified. Of the 5,755 sampled children, 2,173 surveys were returned (37.8%).

Table 1b.4b: Characteristics of children for whom survey responses were received, KPNW study, Top 5 individual providers with highest number of surveys

Characteristics (n=2173)	All Children	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		7 (n=80)	53 (n=77)	4 (n=74)	1 (n=67)	43 (n=66)
Age of child						
Under 10 months of age		22.0%	20.0%	19.5%	20.3%	22.4%
10 to 18 months of age	26.6%	25.0%	29.9%	35.1%	22.4%	15.2%
19-47 months of age	51.4%	55.0%	50.6%	44.6%	55.2%	63.6%
Gender of child						
Female child	46.2%	48.8%	49.4%	47.3%	41.8%	45.5%
Male child	53.8%	51.3%	50.6%	52.7%	58.2%	54.5%
Race/ethnicity of child						
White, non-Hispanic	72.9%	84.8%	77.0%	93.2%	76.9%	62.5%
Asian, non-Hispanic	7.8%	2.5%	6.8%	1.4%	3.1%	20.3%
Hispanic	8.9%	6.3%	12.2%	2.7%	10.8%	10.9%
Other race/ethnicity	10.4%	6.3%	4.1%	2.7%	9.2%	6.3%
Child is the first born in the family	52.1%	52.5%	40.8%	35.1%	54.5%	52.3%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	31.3%	21.5%	24.7%	27.0%	29.7%	26.2%
Education level of mother						
High school or less	12.7%	20.3%	3.9%	14.9%	16.7%	6.2%
More than high school	87.3%	79.7%	96.1%	85.1%	83.3%	93.8%

DISPARITIES

Online PHDS: Variation is observed according to a child's age; race/ethnicity; level of risk for developmental, behavioral, or social delays across all quality measures. Non-Hispanic white children are more likely to meet criteria on the Family Centered Care measure. Children of lower educated mothers are less likely than those with more educated mothers to have high Family Centered Care. Non-CSHCN and children high/medium risk are more likely to receive care met family centered care criteria.

Table 1b.4c. Family centered care by child demographics and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	1309	78.0%
9-18 months	1393	81.4%
19-48 months	1138	80.1%
p values (Pearson chi-square)	-	0.04
Gender		
Male	505	81.2%

Female	497	78.6%		
p values (Pearson chi-square)		-	0.26	
Race/ethnicity				
Hispanic	1530	78.7%		
White non-Hispanic	1779	82.9%		
Black non-Hispanic	71	74.7%		
Asian non-Hispanic	70	66.0%		
Other/Multi race, non-Hispanic	60	76.9%		
p values (Pearson chi-square)		-	<0.0001	
Adult survey responds education level				
Did not complete high school	431	70.2%		
Completed high school or higher education			3264	81.8%
p values (Pearson chi-square)			<0.0001	
CSHCN status				
Non-CSHCN	3461	80.3%		
CSHCN	379	75.8%		
p values (Pearson chi-square)		-	0.02	
At risk for developmental delay (online only)				
Low/No risk	1346	84.1%		
High/Moderate risk	437	74.3%		
p values (Pearson chi-square)		-	<0.0001	

KPNW study: Chi-square test indicate that a child met quality measure criteria differed significantly according to: (1) child's race/ethnicity for four (2) child's birth order, (3) child's developmental and behavioral risk status, (4) respondent education level and (5) parent risk for depression.

Table 1b.4d: Mean number of developmental services care components for which quality care was received and the proportion of children meeting criteria for receiving quality developmental services by characteristics of children and families.

Characteristic of Child or Child's Family	% Meeting AGPE_DB Criteria
Child's Age	
Less than 9 mos.	82.0%
10 to 18 mos.	82.8%
19 to 49 mos.	85.2%
p=0.20	
Child's Gender	
Male Child	85.0%
Female Child	82.6%
P=0.13	
Child's Race	
White, Non-Hispanic	86.6%
Asian, Non-Hispanic	64.0%
Hispanic	80.3%
Other Race, Multiple Race	83.1%
P<0.001	
Birth Order	
Not First Born	86.7%
First Born	81.6%
p=0.002	
Child's Risk for Developmental, Behavioral or Social Delays (Using Parent's Evaluation of Developmental Status)	
Low/No Risk	86.3%
At Risk	78.6%
p<0.001	
Respondent Education	
More than High School	84.9%
High School or Less	78.1%

p=0.004
Respondent's Risk for Depression (Using the Kemper Screener)
No Symptoms of Depression 85.6%
Symptoms of Depression 77.8%
P<0.001

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4**
NA

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

De.7. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
NA

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** CAHMI_Data_Dictionary_Family_Centered_Care.pdf

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.
No

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.
NA

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The numerator measures the number of parents who had a well child visit within the last 12 months and who experienced family centered care in 7 specific areas.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator is the number of survey respondents (parents) answering "Yes, Definitely" to all seven family centered care questions. A 100% positive response is needed for quality for this measure.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator is the number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months who answered the Family Centered Care questions on the Promoting Healthy Development Survey (see Attachment A-2, page 12).

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The denominator is the number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months who answered at least four of the seven Family Centered Care questions on the Promoting Healthy Development Survey (see Attachment A-2, page 12).

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Missing data for the Family Centered Care questions excluded from analysis.

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
Surveys missing four or more of the responses to the Family Centered Care questions are excluded from analysis.Approximately 2.6% of parents who started the Online PHDS did not complete the survey (range 0.0-3.2% for top 5 providers with highest number of surveys; see Testing form, pages 21-23 for more detailed information on missing data).

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*
Although no stratification is required, the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the findings by possible vulnerability, should any individual provide have sufficient data (parent responses) to do so. Potential variables for stratification include:
(1) Child demographic characteristics (e.g., the child´s age, race);
(2) Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or
(3) Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
No risk adjustment or risk stratification
If other:

S.12. **Type of score:**

If other:

S.13. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
The numerator is the sum of survey respondents (parents) answering "Yes definitely" to the Family Centered Care questions. The denominator is the sum of all respondents answering the Family Centered Care questions. Surveys missing four or more responses to this set of questions are excluded from analysis. A score of 100% positive responses to the Family Centered Care questions is required to achieve the quality standard for this measure.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
NA

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Data are collected using the parent-reported "Promoting Healthy Development Survey" (PHDS) developed by the CAHMI (www.wellvisitsurvey.org). Instructions for survey completion are included with the survey. Family Centered Care items are multiple choice (Yes Definitely, Yes, Somewhat, and No, see Attachment A-2, page 12). The PHDS is survey is initiated by the provider who can send it to all parents who have received a well child visit. CAHMI has a website (www.phdstoolkit.org) where providers can register to use for the PHDS. This site assigns each provider a unique URL, which allows for provider identification by CAHMI as well as light branding with the provider´s logo so that it is identifiable by the parent. The PHDS Toolkit website sends an email to the provider with the unique URL link to the survey. The provider then sends the link to the parents asking them with instructions to fill out the survey and provide feedback about the visit. The parent fills out the survey and receives a customized feedback report. The survey data are captured on a secure HIPAA compliant CAHMI server. Through the PHDS Toolkit website, providers can generate a report that aggregate parent data information from the survey. Providers must have a minimum of 10 surveys to generate a report to maintain parent confidentiality. See Evidence Form, Figure 1 for a visual model this process.

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Other

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The Family Centered Care measure is included as part the CAHMI Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org). The data are generated by parents filling out the PHDS. The PHDS is based in English. See Evidence Form, Figure 1 for a description visual model of the data collection process.

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Individual

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Clinician Office/Clinic
If other:

S.22. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
NA

2. Validity – See attached Measure Testing Submission Form

CAHMI_NQF_testing_attachment_Family_Centered_Care_020217.docx

2.1 For maintenance of endorsement

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

Yes

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)

No

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care)

What were the statistical results of the analyses used to select risk factors?

Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

No - This measure is not risk-adjusted

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2965
Measure Title: Family Centered Care
Date of Submission: 2/2/2017
Type of Measure:

☒ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process	☐ Efficiency
☐ Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record

☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Patient reported data	☒ other: Patient reported data

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
We used the following data sources for testing of the quality measure:

1) Online Promoting Healthy Development Survey (PHDS) – data collected through an online, publicly available tool (Promoting Healthy Development Survey-PHDS). Parents who had a well-child care visit in the last 12 months can complete the PHDS. Providers initiate the survey. (See Evidence Form Figure 1 for logic model of the Online PHDS.)
2) Kaiser Permanente Northwest (KPNW) Study – CAHMI partnered with Kaiser Permanente Northwest in Portland, Oregon. The study aimed to evaluate the level and variations in the quality of preventive and developmental services for young children and assess the contribution of key system, provider and patient factors.
3) HRSA Evaluation Study - The specific goal of this study was to evaluate the feasibility, acceptability and impact of three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young children in the context of discussions between pediatric clinicians and parents during well-child visits. The evaluation measures used data from 5 different tools/surveys including PHDS. The parent-completed PHDS was administered before and after the intervention to assess changes in the quality of well-child care. The study funded by Health Resources and Services and Administration's (HRSA) Maternal and Child Health Bureau. (Patient Centered Quality Improvement of Well-Child Care, Final Report, Supported by a grant from the Maternal and Child Health Bureau Research Grants Program, Health Resources and Services Administration, R40 MC08959 03-00.)

1.3. What are the dates of the data used in testing? 2004-2016
Online PHDS: 2008-2016
KPNW Study: 2004-2005
HRSA Evaluation Study: 2010-2012

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Online PHDS: n=5,670 surveys reporting on quality of care provided by 299 individual pediatricians and primary care providers from 88 clinics in 36 states. Participation is a voluntary self-selection process based on knowledge and interest in quality improvement in their practice.

KPNW Study: Provider-level surveys and quality of care assessment were focused on the care provided by 56 individual providers (44 pediatricians, 9 nurse practitioners, 3 physician assistants) in the pediatrics department who were organized into ten geographically distinct offices.

HRSA Evaluation Study: Three pediatric offices in Oregon: 1) a rural site, (4 pediatricians), 2) an urban site (8 pediatricians), and 3) an urban site, (12 pediatricians). All pediatricians in selected clinic and office staff participated in relevant baseline and follow up data collection.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

<u>Online PHDS:</u> Children age 3-48 months of age whose parents completed the online publicly available PHDS were included in the testing. During 2008-2016, we received 5,670 completed surveys. Of those surveys, 5,355 surveys with visited provider IDs were used for analyses. Children's socio-demographic and health characteristics varied across the individual providers included in the analysis.

Table 1.6a: Characteristics of children for whose visited provider ID is available

Characteristics	All Children (n=5355)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		948 (n=91)	1067 (n=90)	1030 (n=77)	802 (n=75)	1022 (n=75)
Age of child						
Under 10 months of age	38.3%	49.5%	33.3%	24.7%	49.3%	34.7%
10 to 18 months of age	34.7%	38.5%	38.9%	57.1%	34.7%	36.0%
19-47 months of age	27.0%	12.1%	27.8%	18.2%	16.0%	29.3%
Race/ethnicity of child						
White, non-Hispanic	53.8%	81.%	20.3%	17.3%	94.4%	53.6%
Hispanic	40.8%	14.3%	74.7%	78.7%	2.8%	37.7
Other race/ethnicity	5.3%	4.8%	7.0%	4.0%	2.8%	8.6%
Respondent education level						
Did not complete high school	12.1%	0	34.1%	15.8%	1.4%	8.1%
Completed high school	88.9%	100%	65.9%	84.2%	98.6%	91.9%
Children who qualify for Children with Special Health Care Needs (CSHCN) Screener criteria						
CSHCN	10.1%	8.8%	10.0%	5.2%	21.3%	14.7%
Non-CSHCN	89.9%	91.2%	90.0%	94.8%	78.7%	85.3%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	22.7%	24.4%	-	0%	25.3%	-

-Data are not available due to small sample size

KPNW Study: The population studied was children 3 to 48 months old who live in a metropolitan area in the Pacific Northwest. One randomly selected child per household whose age would be no younger than 3 months of age and no older than 48 months of age at the time that their parents received the survey and had one or more well-child visits were eligible to be sampled. A random sample of 5,755 children were identified. Of the 5,755 sampled children, 2,173 surveys were returned (37.8%).

Table 1.6b: Characteristics of children for whom survey responses were received, KPNW study, Top 5 individual providers with highest number of surveys

Characteristics	All Children (n=2173)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		7 (n=80)	53 (n=77)	4 (n=74)	1 (n=67)	43 (n=66)
Age of child						
Under 10 months of age	22.0%	20.0%	19.5%	20.3%	22.4%	21.2%
10 to 18 months of age	26.6%	25.0%	29.9%	35.1%	22.4%	15.2%
19-47 months of age	51.4%	55.0%	50.6%	44.6%	55.2%	63.6%
Gender of child						
Female child	46.2%	48.8%	49.4%	47.3%	41.8%	45.5%
Male child	53.8%	51.3%	50.6%	52.7%	58.2%	54.5%
Race/ethnicity of child						
White, non-Hispanic	72.9%	84.8%	77.0%	93.2%	76.9%	62.5%
Asian, non-Hispanic	7.8%	2.5%	6.8%	1.4%	3.1%	20.3%
Hispanic	8.9%	6.3%	12.2%	2.7%	10.8%	10.9%
Other race/ethnicity	10.4%	6.3%	4.1%	2.7%	9.2%	6.3%
Child is the first born in the family	52.1%	52.5%	40.8%	35.1%	54.5%	52.3%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	31.3%	21.5%	24.7%	27.0%	29.7%	26.2%
Education level of mother						
High school or less	12.7%	20.3%	3.9%	14.9%	16.7%	6.2%
More than high school	87.3%	79.7%	96.1%	85.1%	83.3%	93.8%

HRSA Evaluation Study: The study inclusion criteria were used to determine which parents/guardians of children were invited to participate in the interventions and/or evaluation from each participating study site:

- Parent has a well-child visit scheduled at this intervention site for one or more of their children.
- The child is scheduled for their 4-month to 3-year-old well-child visit and, therefore, is between the ages of 4 and 40 months (e.g. 40 month old children could be there for their 3 year well-child visit)
- The parent can read and understand English and is able to complete the intervention and evaluation tools.
- For intervention, the parent was able to access the online version of the Plan My Child's Well-Visit tool and the online evaluation survey.

The analysis includes 551 completed surveys at baseline (2010) and 275 completed surveys at follow-up (2011-12).

Table 1.6c. Sample description for baseline and follow-up PHDS respondents

	Baseline (n=551)	Follow-up (n=275)
Visit type of child for whom survey was completed		
4, 6 or 9-month	38.9%	36.2%
12, 15 or 18-month	33.7%	41.3%
24 or 36-month	27.4%	22.4%

Birth order of child for whom survey was completed		
First child	42.2%	56.6%
Not first child	57.8%	43.4%
Race/ethnicity		
White, non-Hispanic	80.3%	83.5%
Hispanic	8.4%	6.6%
Other/multiple, non-Hispanic	8.6%	6.6%
Asian, non-Hispanic	2.7%	3.3%
Insurance type		
Private or private and public	90.7%	86.7%
Public only (includes Medicaid, Medicare, CHIP and Military)	7.6%	12.1%
Other	0.7%	0.4%
None	0.9%	0.8%

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Online PHDS and KPNW study data were used for reliability testing and stratification analysis. Validity findings are presented from a peer-reviewed publication and online PHDS and KPNW study data. Performance analysis was conducted using the online PHDS, KPNW study and HRSA Evaluation Study data.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Online PHDS: Child's age, sex, race/ethnicity, and respondent (parent) age, race/ethnicity, and education level. The survey does not have a question asks about family income due to complexity of collecting income data by self-reported survey. However, the online PHDS has items assessing the family's economic situation: How much trouble does the family have paying for a) child's health and medical expenses; b) supplies like formula, food, diapers, clothes and shoes; and c) health care for the parent.
KPNW Study: Child's age, sex, race/ethnicity, and education level of mother
HRSA Study: Child's age, race-ethnicity, and insurance type

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Internal consistency: Cronbach's alpha was calculated using the Online PHDS and KPNW data. Cronbach's alpha is the most widely used in health care research when multiple-item measures of a concept or construct are employed. The acceptable values of alpha ranges from 0.70 to 0.95. In addition, factor analysis was performed to investigate the dimensionality of the scale.

The primary aim of the Family Centered Care quality measure is to detect difference between providers on the provision of family centered care to families of young children. Provider level reliability was assessed by inter-unit reliability (IUR) using analysis of variance. IUR can be interpreted as the fraction of the variation among provider scores that is due to real differences, rather than due to chance. If the IUR is higher, the ability of the item or scale measure to discriminate across programs is greater. Scales with reliability coefficients above 0.70 provide adequate precision for use in statistical analysis of unit-level comparisons.[1] As the IUR gets smaller, a larger sample is needed in order to reliably discriminate across programs. In the analysis we included providers with 10 or more completed surveys.

Intra-class correlation (ICC) was calculated using ANOVA, as a ratio of the variance between groups over the total variance. The interpretation of the ICC is as the proportion of relevant variance that is associated with differences among measured objects.[2] Fleiss (1981) and Cicchetti and Sparrow (1981) from the medical group state that ICC range categories are: < 0.40 = poor; 0.40 – 0.59 = fair; 0.60 – 0.74 = good; and > 0.74 = Excellent[3]. Values above about 0.7-0.8 are considered acceptable for applied tests. In the analysis we included providers with 10 or more completed surveys.

1. Nunnally, J. C. Psychometric theory (2nd ed). 1978, New York: McGraw-Hill.
2. McGraw, K. O., and Wong, S. P. Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1996:1(1), 30-46.
3. Cicchetti D.V., and Sparrow, S.S. Developing criteria for establishing the interrater reliability of specific items in a given inventory. American Journal of Mental Deficiency, 1981:86, 127-137.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 2a2.3a. Family-Centered Care: Content, Scoring and Internal Consistency, Online PHDS, all providers and top 5 individual providers with highest number of surveys

What is measured	Scoring	Internal Consistency (Cronbach's Alpha)					
Two multi-part items assess the degree to which care is provided in a family centered manner. Includes respect, understanding specific needs of child and concerns of parent, asking how feeling as a parent, understand family "culture" and talking about resources and issues in the community	Mean score on a multi-item scale	All providers	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
			948 (n=91)	1067 (n=90)	1030 (n=77)	802 (n=75)	1022 (n=75)
		0.88*	0.83*	0.81*	0.85*	0.95*	0.82*

*Met criteria for reliability and internal consistency.

Table 2a2.3b. Family-Centered Care: Content, Scoring and Internal Consistency, KPWN study, all providers and top 5 individual providers with highest number of surveys

What is measured	Scoring	Internal Consistency (Cronbach's Alpha)					
Four multi-part items assess whether general and age specific anticipatory guidance topics are addressed. Includes feeding and nutrition, sleeping and physically caring for child, safety and injury prevention, child growth, development, communication and behavior	Mean score on a multi-item scale	All providers	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
			7 (n=80)	53 (n=77)	4 (n=74)	1 (n=67)	43 (n=66)
		0.81*	0.77*	-	0.90*	0.82*	0.78*

*Met criteria for reliability and internal consistency.

Cronbach's alpha for the Family Centered Care measure is 0.88 (Online PHDS), ranging 0.81-0.95 across providers with highest number of surveys. These findings are consistent with the findings of previous peer-reviewed publications.[4,5]

Inter-unit reliability coefficient for the measure scale is within the recommended threshold (0.73) suggesting that the measure reliably detects difference between providers. Intraclass correlation coefficient for the measure is 0.79, indicating that 78.7% of the variance in the mean of the providers is "true" rather than due to chance.

4. Bethell C, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. Pediatrics. 2001 May;107(5):1084-94.
5. Christina Bethell, PhD, MPH, MBA; Colleen H. Peck Reuland, MS; Neal Halfon, MD, MPH; Edward L. Schor, Measuring the Quality of Preventive and Developmental Services for Young Children: National Estimates and Patterns of Clinicians' Performance. Pediatrics, 2004, 113(6):1973-83

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Psychometric analyses demonstrated that the Family Centered Care quality measure scale have strong internal consistency (Cronbach's alpha ranged 0.77-0.90 across individual providers and two data sources) and reliability detect differences between providers (IUR coefficient 0.73 and ICC 0.79). Two different data sources indicate that the Family Centered Care quality measure provides psychometrically reliable assessment of the provision of nationally recommended well-child care with strong internal consistency.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use

(*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

A standard, multistage process was used to ensure validity of the Family Centered Care items and composite measure:

- Focus groups and in-depth cognitive interviews were conducted throughout the survey development process;
- A review of literature identified through Medline or during key informant interviews; and,
- Three Advisory Groups comprised of pediatricians, family practitioners, consumer representatives, public health experts, and researchers, regularly reviewed and provided input on the identification of quality measurement topics and the development of the PHDS.

A "gold standard" does not exist for determining the criterion validity of patient-reported measures of family centered care. However, to ensure the validity of the PHDS quality measure results, we followed rigorous procedures representing best practices within the field to develop the survey questions. To ensure the content validity of measures of parent experiences, we used qualitative methods, including both focus groups and cognitive interviews, to inform development and evaluation of the family centered care questions. Focus groups with families aimed to identify the aspects of health care quality that are important to parents in the area of preventive care for their children. In-depth cognitive testing of the draft survey items was conducted with 15 families representing a range of racial, income and education groups as well as different types of health insurance coverage, age of child, age and sex of parent, and number of children in family. Focus groups and cognitive interviews with 35 health care providers in Vermont and Washington and 20 parents of young children in Vermont were conducted to inform item-reduction, administration specifications, and reporting templates. Survey modifications were made based on findings in order to improve the reliability, validity and cognitive ease of the FCC items.

Factor analysis was conducted to assess the construct validity of the Family Centered Care quality measure. Each of the survey items used to construct the PHDS scale-based quality measures were used in the factor analysis.[1] Acceptable level of factor loading for instruments developed for research purposes can be as low as 0.60[2] and factor loading more than this threshhold is considered as a strong association.[3] Pearson correlation coefficients were calculated to assess the degree to which each of the item provide unique information.

To assess the concurrent validity of the measure scale, hypothesized associations among Family Centered Care items were examined using logistic regression model (KPNW Study data). We evaluated the hypotheses: Respondents who indicate that providers talked with them about recommended anticipatory guidance topics or providers who discussed family psychosocial issues are more likely to report receiving family-centered care.

Pearson correlation coefficients were calculated across all scale measures to test hypotheses about expected relationships among the Family Centered Care measure and to assess the degree to which this measure provides unique information. We expect a moderate or strong correlation between the family assessment scale measures (>0.30) and inter-scale correlation coefficients to be less than 0.80.

1. Bethell C, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. Pediatrics. 2001 May;107(5):1084-94.
2. Suhr D and Shay M. Guidelines for reliability, confirmatory and exploratory factor analysis. Accessed at: http://www.wuss.org/proceedings09/09WUSSProceedings/papers/anl/ANL-SuhrShay.pdf. Retrieved 02/01/2017
3. Costello A.B and Osborne J.W. Best Practices in Exploratory Factor Analysis: Four

recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation. 2005:10(7). Accessed at: http://www.pareonline.net/pdf/v10n7.pdf, Retrieved 02/01/2017

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Using behavior coding methods, for each item in the FCC quality measure, instances where the respondent required clarification or did not appropriately answer an item were noted. Also, items where the interviewer had difficulty asking the question without edits to the wording were noted. Data analysis was used to inform item-reduction. Content was revised and refined iteratively with each set of interviews.

Cognitive testing confirmed the readability of the FCC items for people across a range of educational levels. Parents were uniformly able to complete the self-administered survey in 10-15 minutes. Readability assessments indicated the PHDS items to be written at the 8th-9th grade reading level. Survey design and formatting was finalized with input from a group of experts and family representatives.

Factor analysis demonstrated a strong factor structure within the FCC quality measures. Each of the items used to construct the FCC quality measure was used in the factor analysis. Average factor loading for FCC was 0.70. Inter-item correlations ranged between 0.55-0.68. Factor analysis suggests that the scale items are unidimensional.

Concurrent validity testing showed that parents who reported that their questions on specific anticipatory guidance topics were answered (odds ratio [OR]: 4.1, 95% confidence interval [CI]: 2.7-5.8, $p<0.001$) or provider discussed about family psychosocial (odds ratio [OR]: 1.01, 95% confidence interval [CI]: 1.0-1.01, $p=0.01$) were more likely to receive family-centered care.

Correlations between the PHDS quality measures were not so high as to suggest redundancy across measures (average correlation: 0.34). As expected, the highest correlation observed was between the "Assessment of family psychosocial well-being" & "Assessment of smoking, drug and alcohol use and safety in the family" (0.54) and "anticipatory guidance from providers" & the "family-centered care" measures (0.52).

Table 2b2.3. Pearson Correlation Coefficients among PHDS Quality Measures (online PHDS)

Scale Measures	Anticipatory Guidance and Parent Education	Family Centered Care	Ask About Parental Concern	Assessment of smoking, drug and alcohol use and safety in the family	Assessment of family psychosocial well-being
Family Centered Care	.52				
Ask About Parental Concern	.16	.14			
Assessment of smoking, drug and alcohol use and safety in the family	.16	.13	.07		
Assessment of family psychosocial well-being	.19	.16	.09	.54	

Average correlation: 0.34

All FCC items have been used in the National Survey of Children's Health. The FCC quality measure is among the few recognized in the Agency for Healthcare Research and Quality's Child Health Toolbox and the National Quality Measures Clearinghouse as measures that meet basic criteria for use as standardized indicators of health care quality for children.

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The FCC quality measure provides conceptually and psychometrically valid assessment of the provision of nationally recommended family centered care to families of young children, with strong construct validity (average factor loading: 0.70). Each of the five composite quality measures provides unique information about performance. Regardless of the population group or the aspect of health care assessed, the quality of health care rarely can be represented accurately by either a single composite performance measure or by assessing whether a single recommended service is provided. The measure is used in national surveys and recognized as measures that meet basic criteria for use as standardized indicators of health care quality for children. The measure serves as an important complement to existing quality measures.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Not applicable

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Not applicable

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Not applicable

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☒ **Stratification by** variable number of **risk categories**

☐ **Other,**

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

Not applicable

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

The Family Centered Care quality measure does not require risk adjustment because we do not expect variation in the quality of family centered care provided for children due to risk factors, e.g. children with special health care needs. The performance should be the same regardless of risk factors. The national experts extensively reviewed the risk

adjustment requirements during development of the Family Centered Care items and did not recommend risk-adjustment for this measure. In addition, during the KPNW study, we did assessment of whether the probability of receiving guidance, education or screening was higher according to a child's level of need or risk, thereby indicating that providers are customizing care to children. The study found no evidence emerged that providers customize care to children most at risk.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Identification of variation in Family Centered Care across subgroups of children helps to highlight aspects of care and population of children for which preventive and developmental services may be most need of improvement. Although no stratification is required (number of surveys for each individual providers may not be sufficient to stratify), the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the quality measures by possible vulnerability:

- Child demographic characteristics (e.g., the child's age, race)
- Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs)
- Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

Based on extensive literature review and expert panel, we identified that child and parent demographics such as age, sex, race-ethnicity, income, insurance, parent behavior, CSHCN screener and follow-up for children at risk can be used for stratification. Several studies have documented differences in access and quality of care provided to children, as well as in parent-reported satisfaction with care.[1-2] One study found that 94.0% of parents reported 1 or more unmet needs, including for Family Centered Care and that there are significant variations in performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors. "[3]

The KPNW study assessed child and family characteristics to characterize the child and their family based on the Family Centered Care item responses: child's race/ethnicity, birth order, risk for developmental, behavioral, or social delays using responses to Frances Glascoe's Parents' Evaluation of Developmental Status (PEDS) items included in the ProPHDS 29 parent's education; and whether he/she is experiencing symptoms of depression using Kathy Kemper's screening items. Adjusted odds ratios were calculated using logistic regression analysis in order to assess differences in the odds of meeting quality measure criteria according to child, family and provider characteristics, after controlling for other variables.

References:

1. Halfon N, Regalado M, Sareen H,Inkelas M, Reuland CH, Glascoe FP, Olson LM. Assessing development in the pediatric office. Pediatrics. 2004 Jun;113(6 Suppl):1926-33.
2. Weech-Maldonado R, Morales LS, Spritzer K, Elliott M, Hays RD. Racial and ethnic differences in parents' assessments of pediatric care in Medicaid managed care. Health Serv Res. 2001 Jul;36(3):575-94.
3. Bethell C, Reuland CH, Halfon N, Schor EL. Measuring the quality of preventive and developmental services for young children: national estimates and patterns of clinicians' performance. Pediatrics. 2004 Jun;113(6 Suppl):1973-83.

2b4.4a. What were the statistical results of the analyses used to select risk factors?
Not applicable

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

See 2b4.3.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)
Pearson's chi-squire test was used to compare the prevalence of FCC quality measure across the stratification characteristics. We preformed logistic regression analysis in order to assess differences in the odds of meeting the Family Centered Care quality measure criteria according to child, family and provider characteristics, after controlling for other variables.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Not applicable

2b4.9. Results of Risk Stratification Analysis:

Online PHDS: Variation is observed according to a child's age; race/ethnicity; level of risk for developmental, behavioral, or social delays across all quality measures. Non-Hispanic white children are more likely to meet criteria on the Family Centered Care measure. Children of lower educated mothers are less likely than those with more educated mothers to receive family centered care. Non-CSHCN and children high/medium risk are more likely to receive care that met family centered care criteria.

Table 2b4.9a. Family centered care by child demographics and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	1309	78.0%
9-18 months	1393	81.4%
19-48 months	1138	80.1%
p values (Pearson chi-square)	-	0.04
Gender		
Male	505	81.2%
Female	497	78.6%
p values (Pearson chi-square)	-	0.26
Race/ethnicity		
Hispanic	1530	78.7%
White non-Hispanic	1779	82.9%
Black non-Hispanic	71	74.7%
Asian non-Hispanic	70	66.0%
Other/Multi race, non-Hispanic	60	76.9%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	431	70.2%

Completed high school or higher education	3264	81.8%
p values (Pearson chi-square)		<0.0001
CSHCN status		
Non-CSHCN	3461	80.3%
CSHCN	379	75.8%
p values (Pearson chi-square)	-	0.02
At risk for developmental delay (online only)		
Low/No risk	1346	84.1%
High/Moderate risk	437	74.3%
p values (Pearson chi-square)	-	<0.0001

KPNW study: Chi square test indicates that a family receiving high quality family centered care differed significantly according to: (1) child's race/ethnicity for four (2) child's birth order, (3) child's developmental and behavioral risk status, (4) respondent education level and (5) parent risk for depression.

Table 2b4.9b: Mean number of developmental services care components for which quality care was received and the proportion of children meeting criteria for receiving quality developmental services by characteristics of children and families.

Characteristic of Child or Child's Family	% Meeting AGPE_DB Criteria
Child's Age	
Less than 9 mos.	82.0%
10 to 18 mos.	82.8%
19 to 49 mos.	85.2%
	p=0.20
Child's Gender	
Male Child	85.0%
Female Child	82.6%
	p=0.13
Child's Race	
White, Non-Hispanic	86.6%
Asian, Non-Hispanic	64.0%
Hispanic	80.3%
Other Race, Multiple Race	83.1%
	P<0.001
Birth Order	
Not First Born	86.7%
First Born	81.6%
	p=0.002
Child's Risk for Developmental, Behavioral or Social Delays (Using Parent's Evaluation of Developmental Status)	
Low/No Risk	86.3%
At Risk	78.6%
	p<0.001
Respondent Education	
More than High School	84.9%
High School or Less	78.1%
	p=0.004
Respondent's Risk for Depression (Using the Kemper Screener)	
No Symptoms of Depression	85.6%
Symptoms of Depression	77.8%
	P<0.001

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

The demographic and socioeconomic survey items included in the Family Centered Care measure make it possible for providers to identify populations and subgroups for which health service delivery improvement is most needed.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)
Not applicable

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Differences in the Family Centered Care quality measure scores across providers is demonstrated for (1) the 5 individual providers with the highest number of surveys using Online PHDS data; (2) across 56 providers using KPNW study data; and (3) pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.

Online PHDS: The performance scale for the Family Centered Care quality measure was calculated using the scoring methods described in Attachment A-4. Individual provider level differences in performance were illustrated by the proportion of children meeting the quality of care criteria across the 5 top providers with the highest number of completed surveys after their well-child visit.

KPNW Study: The significance of differences observed in the proportion of children meeting criteria for the Family Centered Care quality measure across pediatric providers (n=56) was evaluated using t-tests. The relative spread in the FCC score across providers was assessed using the coefficient of variation statistics (standard deviation across providers multiplied by 100%).

HRSA study: Quantitative data results for the baseline (2010) and follow-up (2011-12) study of the intervention sites using the HRSA Evaluation Study data were conducted using basic descriptive statistics to describe each sample and applying chi-square test of statistical significance to assess differences in the quality measure for the baseline and follow-up samples.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Online PHDS: Table 2b5.2a presents the proportion of children whose care met for the quality measure across 5 providers. The proportion of parents who reported receiving family-centered care ranged 78.9%-89.3%.

Table 2b5.2a: Proportion of Children Meeting Measure Criteria, Top 5 individual providers with highest number of surveys

Characteristics	All Children (n=5355)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		948 (n=91)	1067 (n=90)	1030 (n=77)	802 (n=75)	1022 (n=75)
Received family centered care	79.8%	85.7%	80.9%	83.1%	78.9%	89.3%

KPNW Study: 83.9% parents reported that their child's well child care met family-centered care criteria. The variation across providers was statistically significant (p=<0.001).

Table 2b5.2b: Proportion of all children in the study who met criteria for receiving quality services and ranges in proportion across providers. (SD=Standard Deviation)

Developmental Services Quality Measures	Proportion of All Children Meeting Measure Criteria (n = 2173)	Range in the Proportion of Children Meeting Measure Criteria Across 51 Pediatric Providers	Relative Variation (COV) in Measure Scores Across Pediatric Providers
Received family centered care	83.9%	52.8% to 95.8% SD: 9% (p = <0.001)	10.7%

Only providers with n=15 or more PHDS responses are included in the provider level analysis. Provider level n ranges from 15 to 153.

HRSA study

The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the FCC quality of care measures. Parents were more likely to report their needs met for family centered care at the follow-up assessment than at the baseline assessment; and parents were more likely to be asked about one or more psychosocial (family assessment) topics at follow-up. The tables below present comparison of percent of children whose families reported receiving family centered care that met the quality criteria between baseline and follow-up survey data.

Table 2b5.2c. Family Centered Care Measure Comparison by Children's Characteristics
Parent received family-centered care

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	61.9% (343)	65.1% (209)	0.38
10-18 months	64.9% (261)	69.3% (224)	0.23
19-48 months	65.2% (227)	68.0% (151)	0.53
Race			
Hispanic	62.0% (62)	71.1% (32)	0.35
White	64.7% (667)	67.7% (485)	0.20
Asian	39.3% (11)	70.6% (12)	0.07
Multiple or other	66.7% (16)	66.7% (12)	1.00
Insurance type			
Private or private and public	64.9% (704)	67.9% (461)	0.20
Public only (includes Medicaid, Medicare, CHIP, and Military)	58.1% (111)	66.9% (107)	0.10
Other insurance type	(4)	(2)	N/A
Uninsured	66.7% (8)	(4)	N/A
At risk of developmental delay			
Low/no risk	65.5% (669)	69.4% (379)	0.12
High/moderate risk	57.4% (159)	57.9% (99)	0.92

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

Significant gaps and unexplained variations remain in the quality of developmental services for young children. The quality measure assessed here provide a relatively comprehensive picture of performance in the area of preventive and developmental services for young children.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
Not applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

Table 2: Proportion of all children in the study who met criteria for receiving quality developmental services across six components of care and ranges in proportion across providers and offices. (SD=Standard Deviation)
Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The Family Centered Care items were developed based on several rounds of cognitive interviews with parents to ensure quality of responses appropriate to the questions and minimize missing responses.

Online PHDS: Rate of survey completion was calculated based on survey start and complete dates for each respondent. According to the quality measure scoring protocol, if a parent answered less than half of the items in the FCC quality measure, their score is considered to be missing. This does not include items that should have been appropriately skipped. Missing responses are not given a valid score and are not included in the calculation of the quality measure.

KPNW Study: Of the 5,755 sampled children, 2,173 surveys were returned (37.8%). For these children, the provider the parent identified and the provider to which the child was assigned by the health plan were the same 97.3% of the time. A 95% response rate was obtained for the provider survey.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Online PHDS data show that overall 2.6% of parents who started the survey did not complete the survey. Table 2b7.2a presents the frequency of missing values for the Family-Centered Care measure.

Table 2b7.2a. The frequency of missing values for Family-Centered Care measure, overall and top 5 providers

Quality measures	Overall	Provider ID				
		948 (n=91)	1067 (n=90)	1030 (n=77)	802 (n=75)	1022 (n=75)
	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Family Centered Care	3.8 (217)	0	1.1 (1)	0	0	3.2 (5)

KPNW study: Children whose parents responded were not different from those who did not respond in terms of their gender and insurance type. The responding population were slightly less likely to be in the 19 to 48 month age group (55.8% sampled, 51.5% responding) and were somewhat more likely to have had more than one well-visit in the past (67.5% sampled, 74.7% responding).

Table 2b7.1. Sociodemographic Characteristics of KPNW Starting and Responding Sample

Characteristic	Proportion of Starting Sample (N=5755)	Proportion Respondents as of (N=2162)
Gender of Child[NS]		
Male child	52.7	53.7
Female child	47.3	46.3
Age of the Child[S]		
Child age 3-9 months	19.4	21.8
Child age 10-18 months	24.9	26.7
Child age 19-48 months	55.8	51.5
Type of Insurance[NS]		
Private	98.6	98.5
Public	1.4	1.5
Child's Health care utilization		
Number of well-child visits [S]		
1 Well-Child Visit	32.5	25.3
2 or More Well-Child Visits	67.5	74.7
Number of emergency room/urgent care visits		
0 ER/urgent care visits	49.8	51.0
1 ER/urgent care visit	26.2	25.8
2 or more ER/urgent care visit	24.0	23.2
Number of overnight hospital stays [NS]		
0 overnight hospital stays	96.6	96.9
1 or more overnight hospital stays	3.4	3.1

[S]Denotes variables for which statistically significant variation exists between the starting and responding sample for the target child or respondent characteristic.

[NS]No significant variation exists between the starting and responding sample for the target child or respondent characteristic.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Information about non-respondent is not available to compare with those who responded the survey because online PHDS is publicly available tool. However, the low rate of incomplete survey (2.6%) suggests that the measure was acceptable to respondents. Overall, the quality measure had less than 4% of missing cases, ranging 0-3.2% across the top 5 provider with highest number of surveys. Few overall missing values suggest that the measure level results unlikely to be biased by non-response to the survey questions.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Other

If other: Data are generated by parents completing the CAHMI-developed Promoting Healthy Development Survey (PHDS), which is sent to them following a well child visit.

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

Patient/family reported information (may be electronic or paper)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

Data are parent-report using the CAHMI developed Promoting Healthy Development Survey (PHDS). CAHMI captures the data at the provider level through a process described above and in the Evidence Form, Figure 1.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.

During 2012-2016 we experienced some operational delays. In 2012, the provider feedback reports were not automated. When providers wanted a summary report, CAHMI had to manually create them. This was excessively time consuming and CAHMI did not have resources to continue the manual generation of the reports. We sought and received funding to automate the reports. Some difficulties with contractors and staff change-over caused major delays in the project. Then, CAHMI moved from the Oregon Health & Sciences University to Johns Hopkins University School of Public Health in 2014, and it was necessary to upgrade the CAHMI servers. No technical support was available for the transition which caused further delays. Additionally, the PHDS was originally developed in 2001; thus much of the coding and back-end technology for this tool was antiquated and ceased to function after the move. Consequently, and as a result of new improved technology, we have had to redesign the two PHDS related websites - the PHDS toolkit and the parent survey -- as well as the CAHMI PHDS database. Lack of funding caused delays. However, we anticipate launching the new PHDS in February 2017.

Time and cost of data collection are low: provider registration takes about 10 minutes and the parent survey takes about 15-20 minutes to complete. To date, implementation has been limited by lack of funding and resources for outreach, communication and technical support. Our experience in the development and evaluation of the PHDS demonstrated a clear and compelling need to work closely with providers to overcome the many myths that both parents and providers have about patient-engagement quality improvement tools. For the PHDS to be adopted by providers, it is essential to demonstrate, for example, that tool adds value for both the parent and provider, that it fits into and typically improves work flow in the office; improves parent-provider communication, and most important, improve the quality and delivery of nationally recommended services for children. This can only really be accomplished by collaboration and partnership with providers.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Professional Certification or Recognition Program Quality Improvement (Internal to the specific organization)	

4a.1. **For each CURRENT use, checked above (update for maintenance of endorsement), provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

NA

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
The PHDS toolkit (www.phstoolkit.org) and the parent-reported PHDS (www.wellvisitsurvey.org) were used by 68 uniquely identified providers across the country through 2013. We are happy to provide a list of these providers to NQF if desired. In 2014, CAHMI moved from the Oregon Health & Sciences University, Portland OR to the Johns Hopkins University, Baltimore, MD. As a result of the move, and because both server and database technologies had rapidly evolved and improved over the past few years, it was necessary to upgrade our servers, which in turn caused some technical issues with the links between the provider toolkit, the PHDS, and the CAHMI PHDS database. Additionally, the PHDS was originally used to compare providers within a practice as well as between practices within a health system. The anticipated use of the Online PHDS is intended to provide feedback only for individual providers and at the clinic or practice level but not between providers. The combination of these factors led to a decision to upgrade and redesign the PHDS toolkit, PHDS database and Parent Survey. (The PHDS parent survey itself, however, remains fully operational, although use has been nominal from 2014-present, and can be accessed at

www.wellvisitsurvey.org.) The redesign required additional time, IT and CAHMI staff resources and delays were incurred during 2014-2015. However, we are now in the process of finalizing the PHDS Toolkit and database redesign, which is anticipated to be completed and launched in February 2017.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

The redesigned PHDS system (registration toolkit, parent survey tool and PHDS database) is anticipated to be completed and fully functional by February 2017. We have a communication and outreach plan to promote the PHDS as part of the CAHMI Cycle of Engagement (see Attachment A-5), which includes the CAHMI Well Visit Planner (www.wellvisitplanner.org) -- a free parent engagement tool that helps prepare parents for the upcoming well child visit – and the post-visit PHDS which assesses whether the parent received services in alignment with national guidelines as well as family centered care. We have been promoting the Cycle of Engagement in national meetings (AMCHP, PAS, APHA, AcademyHealth ARM, National Child Heath Policy Meeting, and more) over the past several years. We presented the Cycle of Engagement at the CMS Quality Meeting December 13, 2016 and have further plans to unveil the redesigned version at meetings in 2017. The WVP and PHDS have also been endorsed tools that meet requirements for Bright Futures implementation.

We have received substantial interest in the CAHMI parent-engagement tools (both the WVP and the PHDS) from and are in extensive conversations with a number of organizations and agencies including health systems, payers, provider organizations – (CMS/Medicaid, Title V, Head Start, Kaiser Permanente and others); professional associations such as the American Academy of Pediatrics, Bright Futures, National Medicaid Medical Directors, the Children's Hospital Association (CHA), AcademyHealth, Association of Maternal and Child Health Programs (AMCHP), CityMatCH, National Initiative for Children's Healthcare Quality (NICHQ), Autism Speaks, Prevent Child Abuse America; National Prevention Information Network (NIPN); national community-based programs and organizations; philanthropic funders; software platform and electronic medical records systems developers and family organizations. We are in the process of securing funding for Cycle of Engagement EMR integration and implementation projects in partnership with or from a number of interested parties. Further, we are finalizing our application to the American Board of Pediatrics to have the Online PHDS certified as a web-based Maintenance of Certification (MOC) (Part 4) quality improvement (QI) tool for pediatricians. ABP has expressed significant interest in the PHDS and provided some initial funding for the redesign efforts.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Based on PHDS feedback results from an evaluation study conducted in 2011-2012 in Oregon, family centered care did not significantly improve between baseline assessment (69.4%, n=370) and post-intervention assessment (70.1%, n=176, AOR: 1.08, CI: 0.75-1.57), in part because levels of family centered care in this population was already relatively high.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**

There were no unintended or unexpected consequences that we are aware of.

4c.2. **Please explain any unexpected benefits from implementation of this measure.**

There were no unexpected benefits that we are aware of.

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
Extensive qualitative interviews with providers and parents have been conducted and previously reported (See the Evidence Report)

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
Key informant interviews and focus groups with parents and providers were held throughout the testing and evaluation period. We obtained baseline and post-implementation information from providers and post-implementation information from parents. It was necessary to work closely with practices to demonstrate value of the family engagement tools (Well Visit Planner and PHDS) as well as to modify the process to fit individual practice office culture and work flow. A significant amount of provider and staff education was needed to overcome fears and myths that the tool would add to, not help, time management and that parents would not want to participate. This was accomplished by continued and persistent relationship building, spending much time in the office setting with the staff and providers and holding frequent Q&A sessions as the process unfolded.

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.
Through key informant interviews and focus groups with parents and providers.

4d2.2. **Summarize the feedback obtained from those being measured.**
The Family Centered Care measure is seen by providers as an excellent way by which they can improve the quality of the well child visit. In particular this matters a great deal to the providers who are being financially incentivized for family-centered care outcomes.

4d2.3. **Summarize the feedback obtained from other users**
For the most part, parents appreciated being asked about their experience with their well child visits and used it as a way to provide confidential feedback to the providers.

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
The feedback was helpful for future implementation efforts of CAHMI's family engagement tools. The feedback, however, did not result in any changes to the measure itself.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization of Related Measures
The measure specifications are harmonized with related measures;

OR
The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications harmonized to the extent possible?
Yes

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
NA

5b. Competing Measures
The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
NQF #0011 - the PHDS (Promoting Healthy Development Survey) - was endorsed by NQF on October 4, 2012. The PHDS contains the Family Centered Care measure. Neither the questions nor the scoring of the questions have changed since the PHDS was endorsed. It is not actually a competing measure; rather, the Family Centered Care measure is embedded in the PHDS tool.

Please note: The PHDS endorsement (NQF# 0011) and the process measure - Children Receiving Family Centered Care (NQF# 1333) can be found on the NQF measures website but do not appear to be found in the NQF directory in Question 5 above. Hence, we were forced to enter a "no" to Q5 in order to submit this application.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Attachment_A_Supplemental_Materials_Revised_01_18_17-636203528968815669.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Child and Adolescent Health Measurement Initiative
Co.2 **Point of Contact:** Christina, Bethell, cbethell@cahmi.edu, 443-287-5092-
Co.3 **Measure Developer if different from Measure Steward:** Child and Adolescent Health Measurement Initiative
Co.4 **Point of Contact:** Christina, Bethell, cbethell@cahmi.edu, 443-287-5092-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
National Advisors for Patient Centered Quality Improvement of Well-Child Care:
Betsy Anderson, Family Voices
David Bergman, Stanford University
Dimitri Christakis, University of Washington
Paula Duncan, University of Vermont
Cynthia Minkovitz, Johns Hopkins School of Public Health
Amy Perritti, American Academy of Pediatrics
Ed Schor, The Commonwealth Fund

Judy Shaw, University of Vermont Sara Slovin, Johns Hopkins Medicine
Measure Developer/Steward Updates and Ongoing Maintenance Ad.2 **Year the measure was first released:** 2002 Ad.3 **Month and Year of most recent revision:** 01, 2017 Ad.4 **What is your frequency for review/update of this measure?** 3 years Ad.5 **When is the next scheduled review/update for this measure?** 01, 2018
Ad.6 **Copyright statement:** None Ad.7 **Disclaimers:** None
Ad.8 **Additional Information/Comments:** None

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3222
Corresponding Measures:
Measure Title: Assessment of Family Alcohol Use, Substance Abuse and Safety
Measure Steward: Child and Adolescent Health Measurement Initiative
Brief Description of Measure: This measure is used to evaluate the proportion of children whose parents reported being assessed for one or more of the recommended topics regarding alcohol use, substance abuse, safety, and firearms in the home.
Developer Rationale: Recommended developmental services, as set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau, include alcohol and drug use; presence of guns; family violence; and other safety issues in the family. In order to gauge the quality of recommended care provided, this type of information must be collected from the parent in order to identify the level at which providers discuss these issues with parents. Previous studies have shown that parents are willing to discuss such sensitive topics with providers. Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child´s health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews which are not accurate for the specific level of information obtained in the Promoting Health Development Survey (PHDS).

Numerator Statement: The numerator measures the number of parents who had a well child visit within the last 12 months and who were asked about alcohol use, substance abuse, safety and firearms in the house.
Denominator Statement: The denominator is the number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months and answered all of the Family Alcohol Use, Substance Abuse and Safety questions on the Promoting Healthy Development Survey(PHDS, see Attachment A-2, page 17).
Denominator Exclusions: Missing data were excluded from the analysis.

Measure Type: Outcome: PRO
Data Source: Other
Level of Analysis: Clinician : Individual

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary

- This is a Patient-Reported Outcome-Based Performance Measure (PRO-PM) derived from the responses to three questions on the Promoting Healthy Development Survey (complete survey starts on page 20 of the Appendix).
- The developer provided a logic model in both graphic and narrative: (1) the parent and child attend a well child visit with their provider; (2) the provider subsequently sends a survey -- the Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org), which includes one question (3 items, see Attachment A-2, page 17) for the parent to complete; (3) when at least 10 surveys have been completed, the provider receives a feedback report on parents' experiences of the visit and the extent to which they felt they received appropriate and adequate assessment of their family's alcohol use, substance abuse and safety via the CAHMI PHDS Toolkit website (www.phdstoolkit.org); (4) the provider reviews the report and then can engage in a Plan-Do-Study Act (PDSA) quality improvement process to improve their AFAUSAS quality score.
- The developer also notes that, recommended developmental services, as set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau's Bright Futures guidelines, include assessment on alcohol and drug use; presence of guns; family violence; and other safety issues in the family.
- In the Performance Gap section, the developer notes that a HRSA study "found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to be asked about one or more psychosocial (family assessment) topics, including alcohol use, substance abuse, and safety issues, at follow-up." The results are in the testing attachment.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Guidance from the Evidence Algorithm: Patient-reported outcome (Box 1)→ Relationship between PRO and provider action (Box 2)→ Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer reports significant variation across providers who asked parents about all survey items related to substance abuse and firearms in the home: for the online responses to the survey, 7.7%-48.2% of parents of young children reported that their child's pediatric clinician discussed both substance use and firearms in the home, while 24.2-82.4% asked about at least one.
- The results from the Kaiser Permanente Northwest (KPNW) study indicated 53.1% of children had parents reporting that providers discussed at least one alcohol or substance abuse and safety topic.

Disparities

- The online PHDS results showed variation according to a child's age; race/ethnicity (Hispanic children=37.1%, white children=23.9%, black children=32.0%; Asian children=26.1%); level of risk for developmental, behavioral, or social delays (low/no risk=23.0%, high/moderate risk=30.2%). Non-Hispanic white children are less likely to meet criteria on the Family Assessment measures. Children of lower educated mothers and children at high risk for developmental delay are more likely to have high Family Assessment scores.
- After controlling for demographic and health factors, and provider differences, the KPNW study found differences by age, adjusted odds ratio of 0.55 (less than 9 months=63.1%, 10-18 months=55.8%, 19-49 months=47.4%).

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

- This measure will evaluate the proportion of children whose parents report being assessed for alcohol use, substance abuse, or firearms in the home. This information can be used by providers to determine the level at which they discuss these issues with parents. Armed with this information, providers can engage in quality improvement efforts to improve the rate at which they discuss these matters with parents, thereby improving the quality of the preventative healthcare they provide to children.
- What is missing in the evidence review is the evidence for discussing particular issues during the course of a well-child visit; i.e. if these issues are discussed are there better outcomes for the child? What interventions not discussed in the report may lead to better outcomes; i.e. is the discussion alone sufficient or is it assumed that if there are positive findings that specific actions would follow that would reduce risk to the child?
- "This measure is a PRO. The developer discusses that the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau's Bright Futures guidelines, recommended developmental services which include assessment on alcohol and drug use; presence of guns; family violence; and other safety issues in the family. There is a huge amount of variation in care in relation to the pediatricians screening for alcohol and drug use; presence of guns; family violence; and other safety issues in the family. The developers mention that in one study when they provided a training to providers discussing screening for these areas variation in care was reduced.
- Developer cites three studies: online PHDS data, Kaiser Pemanente, and a HRSA study. In the online PHDS study, only 7.7%-48.2% of the parents were assessed for psychosocial topics by their pediatric clinician; in the KPNW study 53.1% of the parents reported that providers discussed such topics. In terms of proof of concept, the HRSA study found that providers psychosocial assessment behavior changed positively after taking a training, suggesting improvements in the quality of care provided to the children. Disparities were examined in all three studies. The online PHDS study found that children with lower educated mothers and developmental delays were morel likely to have higher family assessment scores; the Kaiser study found that scores differed significantly based on child's age. These findings suggest improvements can be made in the area of psychosocial assessments when provided by pediatric providers.
- The developer reports significant variation across providers who asked parents about all survey items related to substance abuse and firearms in the home: for the online responses to the survey, 7.7%-48.2% of parents of young children reported that their child's pediatric clinician discussed both substance use and firearms in the home, while 24.2-82.4% asked about at least one. The results from the Kaiser Permanente Northwest (KPNW) study indicated 53.1% of children had parents reporting that providers discussed at least one alcohol or substance abuse and safety topic. Compliance with screening was measured at the level of the provider.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Other – patient/family reported survey

Specifications:

- Level of analysis: Clinician – individual
- Interpretation of score: Better quality = Higher score
- This is a patient-reported outcome-based performance measure (PRO-PM)
- Numerator: *The numerator measures the number of parents who had a well child visit within the last 12 months and who were asked about alcohol use, substance abuse, safety and firearms in the house.*
- Denominator: *The denominator is the number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months and answered all of the Family Alcohol Use, Substance Abuse and Safety questions on the Promoting Healthy Development Survey (PHDS, see Attachment A-2, page 17).* [Questions are on page 35 of Appendix A.]

- Exclusions: The developer states that "Surveys in which two or more questions from the Family Alcohol Use, Substance Abuse and Safety section of the PHDS were missing were excluded from analysis." [NQF does not consider this an exclusion as it is defining the population of the measure.]
- The developer includes a calculation algorithm.
- The measure is not risk adjusted nor risk stratified, but the developer states that it can be stratified by variables such as child demographics characteristics (e.g., the child´s age, race); child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or Parent health characteristics, if large enough data sets are available.
- The measure does not use sampling.
- This measure relies on a set of questions within the Promoting Healthy Development Survey (page 35 of the Appendix). This online survey is initiated by the provider who sends it to a parent after a well-child visit. Providers must have a minimum of 10 surveys to generate a report to maintain parent confidentiality.

Questions for the Committee:

- o *Are all the data elements (question items) clearly defined?*
- o *Is the logic or calculation algorithm clear?*
- o *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

NQF Note: Both measure score and data element reliability testing are required for PRO-PMs.

Method(s) of reliability testing

- The developer used data from the online Promoting Healthy Development Survey (PHDS), the KPNW study, and a HRSA evaluation study that tested "three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young children in the context of discussions between pediatric clinicians and parents during well-child visits"; the HRSA study used 5 tools, including the PHDS.
- The Cronbach alpha to test internal consistency (data/item element reliability) was calculated using the Online PHDS and KPNW data. In addition, factor analysis was performed to investigate the dimensionality of the scale.
- Score-level reliability was assessed by inter-unit reliability (IUR) using analysis of variance. Providers with 10 or more surveys were assessed; no information on the N is provided for these analyses.
- Intra-class correlation (ICC) was calculated using ANOVA.

Results of reliability testing

- Using the top 5 individual providers with the highest number of surveys (N=77 to 94) from the online PHDS testing, the developer reports the Cronbach's alpha for internal consistency (item-level) range from 0.69-0.82 with the mean score for all providers at 0.81. However, more details were not provided.
- The developer reports that the results for the inter-unit reliability (IUR) testing are within the recommended threshold (0.71) to reliably demonstrate differences between providers.
    - o [*NQF note: IUR measures the proportion of the measure variability that is attributable to the between facility variance. A small IUR (near 0) reveals that most of the variation of the measures between facilities is driven by random noise, indicating the measure would not be a good characterization of the differences among facilities, whereas a large IUR (near 1) indicates that most of the variation between facilities is due to the real difference between facilities. The recommended range is above 0.70.]*

- The ICC was 0.78; according to the literature cited by the developer values above 0.74 are considered excellent.

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance among clinicians can be identified?*

Guidance from the Reliability Algorithm: Precise specifications (Box 1)→ Empirical reliability testing (Box 2) → Score level testing (Box 4) → Appropriate method used (Box 5) → High certainty or confidence that the performance measure scores are reliable (Box 6a)→ High.

Highest possible rating is HIGH.

Note: PRO-PMS *require* element-level testing as well, which was conducted. If judged without score-level testing, the highest eligible rating for this type of testing is MODERATE.

Preliminary rating for reliability: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**
Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

NQF Note: Both measure score and data element validity testing are required for PRO-PMs.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

The developer conducted several methods of validity testing:

- Factor analysis was conducted to assess the construct validity of the quality measure. A Scree test was used to determine the number of factors to extract. Both oblique and orthogonal rotations were evaluated with promax and varimax methods used, respectively.
- To assess the concurrent validity of the quality measure, hypothesized associations among PHDS items and scales were examined. The developer "tested a hypothesis that respondents who indicate that providers talked with them about keeping house and car safe topics more likely to report increased confidence as a parent because of interactions with health care providers compared with respondents who indicate that providers did not talk with them."

- Pearson correlation coefficients were calculated across all scale measures to test hypotheses about expected relationships among the PHDS quality measures and to assess the degree to which each of the PHDS quality measures provide unique information. The developer notes that "We expect a moderate or strong correlation between the family assessment scale measures (>0.30) and inter-scale correlation coefficients to be less than 0.80."
- For face validity, the PHDS items were tested using focus groups, in-depth cognitive interviews, a literature review, and an advisory board of expert stakeholders.

Validity testing results:
The developer reports the following results:

- "Average factor loading for AFAUSAS was 0.66. Inter-item correlation ranged between 0.54-0.60. Factor analysis suggests that the scale items are unidimensional." Per the developer, acceptable ranges for factor loading are above 0.60.
- The concurrent validity testing results demonstrated improved confidence in protecting children from injury if providers talked about keeping homes and cars safe (odds ratio [OR]: 5.9, 95% confidence interval [CI]: 3.4-10.2; OR: 8.3, 95% CI: 5-13.8) but also showed that parents report they are rarely asked about other psychosocial issues, including gun safety or how parenting works into their daily activities.
- The developer provides a table of Pearson Correlation Coefficients, which assesses whether the measures are examining different topics. The results suggest, according to the developer, that the measures are not redundant, with an average correlation of 0.34. This measure was most highly associated with the *Assessment of family psychosocial well-being measure* (0.54).

Questions for the Committee:

- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:

N/A

Question for the Committee:

- *Is the lack of exclusions consistent with the evidence?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- The developer does not risk adjust the measure because "we do not expect variation in the quality of care provided for children due to risk factors, e.g. children with special health care needs. The provider's performance should be the same regardless of risk factors."
- The developer notes the measure can be stratified by several demographic or health variables as "Identification of variation in quality measures across subgroups of children helps to highlight aspects of care and population of children for which preventive and developmental services may be most need of improvement."
- The developer reports that many studies have shown differences in access to and quality of care, as well as parent satisfaction. The developer states that "One study found: Overall, 94.0% of parents reported 1 or more unmet needs for a number of aspects of care, including assessing family alcohol use, substance abuse and safety. Uninsured children and children aged 18 to 35 months are disproportionately represented among the 15.3% of children whose parents indicated an unmet need this area of care. There are significant variations in performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors."
- Variations were observed by demographic and socioeconomic factors.

Questions for the Committee:

- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To assess meaningful differences, the developer analyzed the top 5 providers (number of individual surveys completed) in the online PHDS; reported on 56 providers using KPNW study data; and reported on pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.
  - Online PHDS: For the top 5 individual providers with the highest numbers of surveys (n=77 to 94), a range of 7.7%-48.2% of parents of young children reported that their child's pediatric clinician discussed all psychosocial topics including alcohol use, substance abuse and safety; the average for all children was 29.6%. Asking about at least of the topics of substance abuse or firearms in the home averaged 60.5% and ranged from 24.2%-82.4%.
  - KPNW Study: 53.1% of children had parents reporting that providers discussed at least one alcohol or substance abuse and safety topic. Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant, 32.4%-69.8% (p=0.04). Provider n ranged from 15-153.
  - When the provider was used as the level 2 clustering variable, only 1.1-2.2% of the total variance observed was explained by either measured or unmeasured differences between providers. The developer indicates that this "suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers" and that "the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across the all quality measures" (i.e., providers are inconsistent and going to a different provider may not improve a child's care). However, the HRSA study does demonstrate that providers can improve their performance with an intervention.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

The developer reports the following:

- **Online PHDS:** Rate of survey completion was calculated based on survey start and complete dates for each respondent. According to the quality measure scoring protocol, if a parent answered less than half of the items in the AFAUSAS measure, his/her score is considered to be missing. This does not include items that should have been appropriately skipped. Missing responses are not given a valid score and are not included in the calculation of the quality measure.
- **Online PHDS** data show that 2.5% of parents who started the survey did not complete the survey.
- **KPNW Study:** Of the 5,755 sampled children, 2,173 surveys were returned (37.8%). For these children, the provider the parent identified and the provider to which the child was assigned by the health plan were the same 97.3% of the time. A 95% response rate was obtained for the provider survey.
- The developer notes that responses for the KPNW survey did not differ by gender or insurance type, but did differ by age and by number of previous well visits.
- The specifications indicate that surveys in which two or more questions from the Family Alcohol Use, Substance Abuse and Safety section of the PHDS were missing were excluded from analysis; however, no information was provided on why this does not bias the responses.

- The developer states that information about non-respondents is not available, but "Overall, the quality measure had 2.5% of missing cases, ranging 0-5.6% across the top 5 providers with highest number of surveys. Few overall missing values suggest that the measure level results unlikely to be biased by non-response to the survey questions."

Guidance from the Validity Algorithm: Specifications consistent with evidence (Box 1)→ Threats to validity addressed (Box 2) →Empirical validity testing (Box 3)→ Measure score testing (Box 6)→ Appropriate method (Box 7)→ Moderate certainty or confidence that the performance measure scores are a valid indicator of quality (Box 8b)→ Moderate

The highest possible score is MODERATE.

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
RATIONALE: Missing data are not fully addressed; non-respondent bias not available

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

- All data elements are clearly defined. The calculation algorithm described the target population: parents; exclusions: surveys missing one or more responses to the set of questions; to achieve quality on this measure, a score of 100% is required (asking all questions). No risk adjustment was made for this measure. Because this is an online survey, it can be easy to consistently implement, with those having access to computers. Families that do not have access to computers, or do not have access to online services, will likely not be included in the results.
- Data elements are clearly defined. The measure is not risk adjusted nor risk stratified. However, the developer states that it can be stratified by variables such as child demographics characteristics (e.g., the child´s age, race); child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or Parent health characteristics, if large enough data sets are available.
- Data was used from all three studies to examine the reliability of the Assessment of Family Alcohol Use and Substance Abuse and Safety measures. Reliability statics were strong. Overall Cronbach's alpha was .81, which was consistent with published studies. The IUR was .71, and the ICC was .78, all within an acceptable range. These results evidence enough sensitivity to detect differences between clinicians.
- Using the top 5 individual providers with the highest number of surveys (N=77 to 94) from the online PHDS testing, the developer reports the Cronbach's alpha for internal consistency (item-level) range from 0.69-0.82 with the mean score for all providers at 0.81. However, more details were not provided. The developer reports that the results for the inter-unit reliability (IUR) testing are within the recommended threshold (0.71) to reliably demonstrate differences between providers. The ICC was 0.78; according to the literature cited by the developer values above 0.74 are considered excellent.
- Factor analysis was conducted to assess the construct validity of the quality measure.
- To assess the concurrent validity of the quality measure, hypothesized associations among PHDS items and scales were examined. The developer "tested a hypothesis that respondents who indicate that providers talked with them about keeping house and car safe topics more likely to report increased confidence as a parent because of interactions with health care providers compared with respondents who indicate that providers did not talk with them."
- Pearson correlation coefficients were calculated across all scale measures to test hypotheses about expected relationships among the PHDS quality measures and to assess the degree to which each of the PHDS quality measures provide unique information. The developer notes that "We expect a moderate or strong correlation between the family assessment scale measures (>0.30) and inter-scale correlation coefficients to be less than 0.80.
- For face validity, the PHDS items were tested using focus groups, in-depth cognitive interviews, a literature review, and an advisory board of expert stakeholders. Average factor loading for AFAUSAS was 0.66. Inter-item correlation ranged between 0.54-0.60. Factor analysis suggests that the scale items are unidimensional." Per the developer, acceptable ranges for factor loading are above 0.60.
- The concurrent validity testing results demonstrated improved confidence in protecting children from injury if providers talked about keeping homes and cars safe (odds ratio [OR]: 5.9, 95% confidence interval [CI]: 3.4-10.2; OR: 8.3, 95% CI: 5-13.8) but also showed that parents report they are rarely asked about other psychosocial issues, including gun safety or how parenting works into their daily activities.

- Overall missing data for the online PHDS survey was low at 2.5%, and does not appear to constitute a threat to validity.
    - The developer does not risk adjust the measure because "we do not expect variation in the quality of care provided for children due to risk factors, e.g. children with special health care needs. The provider's performance should be the same regardless of risk factors."
    - The developer notes the measure can be stratified by several demographic or health variables as "Identification of variation in quality measures across subgroups of children helps to highlight aspects of care and population of children for which preventive and developmental services may be most need of improvement."
    - The developer reports that many studies have shown differences in access to and quality of care, as well as parent satisfaction. The developer states that "One study found: Overall, 94.0% of parents reported 1 or more unmet needs for a number of aspects of care, including assessing family alcohol use, substance abuse and safety. Uninsured children and children aged 18 to 35 months are disproportionately represented among the 15.3% of children whose parents indicated an unmet need this area of care. There are significant variations in performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors."
    - To assess meaningful differences, the developer analyzed the top 5 providers (number of individual surveys completed) in the online PHDS; reported on 56 providers using KPNW study data; and reported on pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.
    - Online PHDS: For the top 5 individual providers with the highest numbers of surveys (n=77 to 94), a range of 7.7%-48.2% of parents of young children reported that their child's pediatric clinician discussed all psychosocial topics including alcohol use, substance abuse and safety; the average for all children was 29.6%. Asking about at least of the topics of substance abuse or firearms in the home averaged 60.5% and ranged from 24.2%-82.4%.
    - KPNW Study: 53.1% of children had parents reporting that providers discussed at least one alcohol or substance abuse and safety topic. Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant, 32.4%-69.8% (p=0.04). Provider n ranged from 15-153.
    - When the provider was used as the level 2 clustering variable, only 1.1-2.2% of the total variance observed was explained by either measured or unmeasured differences between providers. The developer indicates that this "suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers" and that "the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across the all quality measures" (i.e., providers are inconsistent and going to a different provider may not improve a child's care). However, the HRSA study does demonstrate that providers can improve their performance with an intervention.

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- This is a patient reported outcome. Data are generated by parents completing the CAHMI-developed Promoting Healthy Development Survey (PHDS), which is sent to them by their provider following a well child visit.
- Although the survey has been in use since 2001, there is not currently an automated reporting system for providers. The developer has been working on a new website for the survey that will automatically report data, and expects it to launch in February 2017.
- The developer reports that the provider registration takes about 10 minutes and the parent survey takes about 15-20 minutes. There are no fees, licensing requirements, etc., to use the measure.

Questions for the Committee:

- *Is the data collection strategy ready to be put into operational use?*
- *Does the developer have a status update on the new website?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 3: Feasibility

- The developer is working on a new website that will automatically report the data, with a launch data of February 2017. The survey itself takes 15-20 minutes to complete. There are no fees associated with the measure. Again, this will only work for families that have easy online access.
- This is a patient reported outcome. Data are generated by parents completing the CAHMI-developed Promoting Healthy Development Survey (PHDS), which is sent to them by their provider following a well child visit.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No** ☐ **UNCLEAR**
OR
Planned use in an accountability program? ☐ **Yes** ☒ **No**

Accountability program details
No confirmed use for an accountability program, but the developer has been in discussion with a number of organizations that are interested in using the measure, including CMS/Medicaid, Title V, and Head Start.

Improvement results
The developer provided the following response: "Based on PHDS feedback results from an evaluation study conducted in 2011-2012 in Oregon, Family assessment for one or more family assessment topics increased 103.3%, a statistically significant increase, from 21.5% at baseline (2010, n=116) to 43.7% post-assessment (2011-12, n=111, AOR: 3.32, 95% CI: 2.24-4.91)."

Unexpected findings (positive or negative) during implementation
The developer was not aware of any unintended consequences.

Potential harms
The developer was unaware of any potential harms.

Vetting of the measure
N/A

Feedback:
N/A

***Questions for the Committee**:*
- o *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 4: Usability and Use

- The measure is currently not publically reported, or used in any accountability programs. CMS/Medicaid, Title V, and Head Start have all mentioned interest in using this measure. There appear to be no harms associated with using this measure.
- This seems to be a very complex method of assessing a physician's skills at addressing particular issues during the course of a well child visit.
- Not available publicly. No confirmed use for an accountability program, but the developer has been in discussion with a number of organizations that are interested in using the measure, including CMS/Medicaid, Title V, and Head Start. The developer provided the following response: "Based on PHDS feedback results from an evaluation study conducted in 2011-2012 in Oregon, Family assessment for one or more family assessment topics increased 103.3%, a statistically significant increase, from 21.5% at baseline (2010, n=116) to 43.7% post-assessment (2011-12, n=111, AOR: 3.32, 95% CI: 2.24-4.91).

Criterion 5: Related and Competing Measures

Related or competing measures

This measure is part of a set of five based on the PHD survey.

- 3219: Anticipatory Guidance and Parental Education
- 3220: Ask About Parental Concerns
- 3221: Family Centered Care
- 3222: Assessment of Family Alcohol Use, Substance Abuse and Safety
- 3223: Assessment of Family Psychosocial Screening

Harmonization

N/A

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

RATIONALE IF NOT ELIGIBLE: The measure has not been vetted by those being measured or other users.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2968
Measure Title: Assessment of Family Alcohol Use, Substance Abuse and Safety
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: 1/13/2017

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome
☒ Health outcome: Click here to name the health outcome
☒Patient-reported outcome (PRO): Assessment of Family Alcohol Use, Substance Abuse and Safety (AFAUSAS)

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Figure 1 (attached) shows the logic model by which the Assessment of Family Alcohol Use, Substance Abuse and Safety (AFAUSAS) quality measure is obtained and improved. Simply said: (1) the parent and child attend a well child visit with their provider; (2) the provider subsequently sends a survey -- the Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org), which includes one question (3 items, see Attachment A-2, page 17) for the parent to complete; (3) when at least ten surveys have been completed, the provider receives a feedback report on parents' experiences of the visit and the extent to which they felt they received appropriate and adequate assessment of their family's alcohol use, substance abuse and safety via the CAHMI PHDS Toolkit website (www.phdstoolkit.org); (4) the provider reviews the report and then can engage in a *Plan-Do-Study Act* (PDSA) quality improvement process to improve their AFAUSAS quality score. THE PDSA cycle involves reviewing the baseline data; developing and implementing a plan of action to improve the score; obtaining further data from the parent; and comparing the first set of results with the second. The full process is repeated until providers are satisfied with their improved scores. We are currently applying for this process to be approved by the American Board of Pediatrics (ABP) for maintenance of certification (MOC, Part 4) credit. The provider must complete three PDSA cycles. Each time point must have at least 25 completed surveys and there must be at least 8 weeks between time periods.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

Recommended developmental services, as set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau, include alcohol and drug use; presence of guns; family violence; and other safety issues in the family.[1] In order to gauge the quality of recommended care provided, this type of information must be collected from the parent in order to identify the level at which providers discuss these issues with parents. Previous studies have shown that parents are willing to discuss such sensitive topics with providers. Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child's health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews which are not accurate for the specific level of information obtained in the Promoting Health Development Survey (PHDS, see Attachment A-2, page 17).The process outlined in the logic model (1a.12) allows health care providers to better understand the extent to which their patients experience "quality care" – in this case, the extent to which parents received an assessment of family alcohol use, substance abuse and safety. It also allows providers to engage in quality improvement activities to improve their parent-reported Family Centered Care quality score by using several Plan-Do-Study Act (PDSA) cycles, as described above.

[1] *Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents*, 3rd Edition.

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

Figure_1_Family_AUSAS_Logic_Model.docx,CAHMI_AFAUSAS_evidence_attachment_revised_02_02_17_revised.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

No

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

IF a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)

IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.

Recommended developmental services, as set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau, include alcohol and drug use; presence of guns; family violence; and other safety issues in the family. In order to gauge the quality of recommended care provided, this type of information must be collected from the parent in order to identify the level at which providers discuss these issues with parents. Previous studies have shown that parents are willing to discuss such sensitive topics with providers. Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child´s health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews which are not accurate for the specific level of information obtained in the Promoting Health Development Survey (PHDS).

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis***. (This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

see also Testing Form_Assessment Alcohol Drug Safety

DATA SOURCES:

Differences in the Family AUSAS quality measure scores across providers is demonstrated for (1) 5 top individual providers with the highest number of surveys using Online PHDS data; (2) across 56 providers using Kaiser Permanente NW study data; and (3) pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.

Online PHDS: The performance scale for the quality measure was calculated using the scoring methods described in Attachment A-4. Individual provider level differences in performance were illustrated by the proportion of children meeting the quality of care criteria across 5 top providers with the highest number of completed surveys after their well-child visit.

KPNW Study: The significance of differences observed in the proportion of children meeting criteria for the quality measure across pediatric providers (n=56) was evaluated using t-tests. The relative spread in the quality measure score across providers was assessed using the coefficient of variation statistics (standard deviation across providers multiplied by 100%). Multi-level regression models were conducted using the pediatric provider as the level 2 clustering variable, in order to assess the degree to which the probability that a child meets criteria on each quality measure is explained by differences between providers (called the

"clustering effect"). In implementing this multi-level regression method (Empty Model), the presence of a significant clustering effect by pediatric providers was estimated prior to accounting for the child and family characteristics associated with each provider. Second, variables related to the child and family characteristics (child's age, gender, race/ethnicity, birth order, developmental and behavioral delay risk status; parent education and risk for depression) were added to the Empty Model to assess how much of the provider clustering effect observed remains after accounting for these characteristics (called the "Patient Model").

HRSA study: Quantitative data results for the baseline (2010) and follow-up (2011-12) study of the intervention sites using the HRSA Evaluation Study data were conducted using basic descriptive statistics to describe each sample and applying chi-square test of statistical significance to assess differences in the quality measure for the baseline and follow-up samples.

PERFORMANCE RESULTS

Online PHDS: Table 1b.2a present the proportion of children whose care met for the quality measure across 5 providers. Variation across providers who asked parents about all survey items related to substance abuse and firearms in the home is substantially wide across observed providers. Only 7.7%-48.2% of parents of young children reported that their child's pediatric clinician discussed psychosocial topics including alcohol use, substance abuse and safety.

Table 1b.2a: Proportion of Children Meeting Measure Criteria, Top 5 individual providers with highest number of surveys

Characteristics (n=5355)	All Children				
	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
	1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
Asking parents about substance abuse and firearms in the home (at least one of these topics)	60.5%	59.8%	24.2%	82.4%	70.5% 77.0%
Asking parents about substance abuse and firearms in the home (asked all items)	29.6%	29.3%	7.7%	48.2%	38.5% 31.1%

KPNW Study: 53.1% of children had parents reporting that providers discussed at least one alcohol or substance abuse and safety topic. Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant (p=0.04)

Table 1b.2b: Proportion of all children in the study who met criteria for receiving quality services and ranges in proportion across providers. (SD=Standard Deviation)

Developmental Services Quality Measures (n = 2173)	Proportion of All Children Meeting Measure Criteria	Range in the Proportion of Children Meeting Measure Criteria Across 51 Pediatric Providers	Relative Variation (COV) in Measure Scores Across Pediatric Providers
Asking parents about substance abuse and firearms in the home (at least one of these topics)	53.1%	32.4% to 69.8%	SD: 9%; (p = 0.04) 17.9%

Only providers with n=15 or more PHDS responses are included in the provider level analysis. Provider level n ranges from 15 to 153.

Multi-level analysis: For the Empty Model that used the provider as the level 2 clustering variable, only 1.1% to 2.2% of the total variance observed in whether children met criteria for each of the all quality measures was explained by either measured or unmeasured differences between the providers that they see. This suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers. These findings translate into a 1.19 to 1.29 median odds ratio across the six quality measures, indicating that the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across the all quality measures. When child/family level characteristics are added to the model (Patient Model), the total variance explained by differences between providers does not change significantly.

HRSA study

The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to be asked about one or more psychosocial (family assessment) topics, including alcohol use, substance abuse and safety issues, at follow-up. The tables below present comparison of percent of children who received care met the quality care criteria between baseline and follow-up survey data.

Table 1b.2c. Family Assessment*, by Children's Characteristics: ---
Parent was asked about one or more family assessment topics

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	23.9% (132)	45.0% (145)	<0.0001
10-18 months	21.5% (87)	34.4% (111)	<0.0001
19-48 months	29.3% (103)	50.5% (112)	<0.0001
Race			
Hispanic	26.0% (26)	47.8% (22)	0.01
White	24.5% (252)	42.7% (305)	<0.0001
Asian	28.6% (8)	(4)	-
Multiple or other	50.0% (12)	50.0% (9)	1.00
Insurance type			
Private or private and public	22.9% (248)	37.9% (257)	<0.0001
Public only (includes Medicaid, Medicare, CHIP, and Military)	33.5% (64)	39.1% (63)	<0.0001
Other insurance type	(2)	(1)	N/A
Uninsured	(4)	(3)	N/A
At risk of developmental delay			
Low/no risk	24.3% (248)	41.6% (227)	<0.0001
High/moderate risk	26.0% (73)	44.7% (76)	<0.0001

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
NA

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*
See also Testing Form_Assessment Alcohol Drug Safety
DATA SOURCES
We used the following data sources for testing of the quality measure:
(1) Online Promoting Healthy Development Survey (PHDS) – data collected through an online, publicly available tool (Promoting Healthy Development Survey-PHDS). Parents who had a well-child care visit in the last 12 months can complete the PHDS. Providers initiate the survey. (See Evidence Form Figure 1 for visual model of the Online PHDS.)
2) Kaiser Permanente Northwest (KPNW) Study – CAHMI partnered with Kaiser Permanente Northwest in Portland, Oregon. The study aimed to evaluate the level and variations in the quality of preventive and developmental services for young children and assess the contribution of key system, provider and patient factors.

STUDY POPULATION
Online PHDS: Children age 3-48 months of age whose parents completed the online publicly available PHDS were included in the testing. During 2008-2016, we received 5,670 completed surveys. Of those surveys, 5,355 surveys with provider IDs were used for analyses. Children's socio-demographic and health characteristics varied across the individual providers included in the analysis.

Table 1b.4a: Characteristics of children for whose visited provider ID is available

Characteristics (n=5355)	All Children	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)					
		1029 (n=94)		948 (n=91)		1067 (n=90)	
		927 (n=79)		1030 (n=77)			
Age of child							
Under 10 months of age	38.3%	19.1%	49.5%	33.3%	54.4%	24.7%	
10 to 18 months of age	34.7%	39.4%	38.5%	38.9%	29.1%	57.1%	
19-47 months of age	27.0%	41.5%	12.1%	27.8%	16.5%	18.2%	
Race/ethnicity of child							
White, non-Hispanic	53.8%	13.3%	81.0%	20.3%	50.7%	17.3%	
Hispanic	40.8%	81.1%	14.3%	74.7%	40.8%	78.7%	
Other race/ethnicity	5.3%	5.5%	4.7%	7.0%	8.4%	4.0%	
Respondent education level							
Did not complete high school	12.1%	23.6%	0	34.1%	6.4%	15.8%	
Completed high school	88.9%	76.4%	100%	65.9%	93.6%	84.2%	
Children who qualify for Children with Special Health Care Needs (CSHCN) Screener criteria							
CSHCN	10.1%	7.4%	8.8%	10.0%	11.4%	5.2%	
Non-CSHCN	89.9%	92.6%	91.2%	90.0%	88.6%	94.8%	
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	22.7%	-	24.4%	-	28.9%	0%	

-Data is not available due to small sample size

KPNW Study: The population studied was children 3 to 48 months old who live in a metropolitan area in the Pacific Northwest. One randomly selected child per household whose age would be no younger than 3 months of age and no older than 48 months of age at the time that their parents received the survey and had one or more well-child visits were eligible to be sampled. A random sample of 5,755 children were identified. Of the 5,755 sampled children, 2,173 surveys were returned (37.8%).

Table 1b.4b: Characteristics of children for whom survey responses were received, KPNW study, Top 5 individual providers with highest number of surveys

Characteristics (n=2173)	All Children	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)					
		7 (n=80)	53 (n=77)		4 (n=74)	1 (n=67)	43 (n=66)
Age of child							
Under 10 months of age	22.0%	20.0%	19.5%	20.3%	22.4%	21.2%	
10 to 18 months of age	26.6%	25.0%	29.9%	35.1%	22.4%	15.2%	
19-47 months of age	51.4%	55.0%	50.6%	44.6%	55.2%	63.6%	
Gender of child							
Female child	46.2%	48.8%	49.4%	47.3%	41.8%	45.5%	
Male child	53.8%	51.3%	50.6%	52.7%	58.2%	54.5%	
Race/ethnicity of child							
White, non-Hispanic	72.9%	84.8%	77.0%	93.2%	76.9%	62.5%	
Asian, non-Hispanic	7.8%	2.5%	6.8%	1.4%	3.1%	20.3%	
Hispanic	8.9%	6.3%	12.2%	2.7%	10.8%	10.9%	
Other race/ethnicity	10.4%	6.3%	4.1%	2.7%	9.2%	6.3%	
Child is the first born in the family	52.1%	52.5%	40.8%	35.1%	54.5%	52.3%	
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	31.3%	21.5%	24.7%	27.0%	29.7%	26.2%	
Education level of mother							
High school or less	12.7%	20.3%	3.9%	14.9%	16.7%	6.2%	
More than high school	87.3%	79.7%	96.1%	85.1%	83.3%	93.8%	

HRSA Evaluation Study: The study inclusion criteria were used to determine which parents/guardians of children were invited to participate in the interventions and/or evaluation from each participating study site:

- Parent has a well-child visit scheduled at this intervention site for one or more of their children.

- The child is scheduled for their 4-month to 3-year-old well-child visit and, therefore, is between the ages of 4 and 40 months (e.g. 40 month old children could be there for their 3 year well-child visit)
- The parent can read and understand English and is able to complete the intervention and evaluation tools.
- For intervention, the parent was able to access the online version of the Plan My Child's Well-Visit tool and the online evaluation survey.

The analysis includes 551 completed surveys at baseline (2010) and 275 completed surveys at follow-up (2011-12)

Table 1c. Sample description for baseline and follow-up PHDS respondents

	Baseline (n=551)	Follow-up (n=275)
Visit type of child for whom survey was completed		
4, 6 or 9-month	38.9%	36.2%
12, 15 or 18-month	33.7%	41.3%
24 or 36-month	27.4%	22.4%
Birth order of child for whom survey was completed		
First child	42.2%	56.6%
Not first child	57.8%	43.4%
Race/ethnicity		
White, non-Hispanic	80.3%	83.5%
Hispanic	8.4%	6.6%
Other/multiple, non-Hispanic	8.6%	6.6%
Asian, non-Hispanic	2.7%	3.3%
Insurance type		
Private or private and public	90.7%	86.7%
Public only (includes Medicaid, Medicare, CHIP and Military)	7.6%	12.1%
Other	0.7%	0.4%
None	0.9%	0.8%

DISPARITIES

Online PHDS: Variation is observed according to a child's age; race/ethnicity; level of risk for developmental, behavioral, or social delays. Non-Hispanic white children are less likely to meet criteria on the Family Assessment measures. Children of lower educated mothers and children at high risk for developmental delay are more likely to have high Family Assessment scores.

Table 1b.4c. Assessment of smoking, drug and alcohol use and safety (asked about all items) in the family by child demographics and other characteristics

Characteristics	All children n	%
Age groups		
3-8 months	705	31.3%
9-18 months	531	28.1%
19-48 months	403	29.1%
p values (Pearson chi-square)	-	0.06
Gender		
Male	215	34.7%
Female	210	33.6%
p values (Pearson chi-square)	-	0.67
Race/ethnicity		
Hispanic	723	37.1%
White non-Hispanic	680	23.9%
Black non-Hispanic	32	32.0%
Asian non-Hispanic	29	26.1%
Other/Multi race, non-Hispanic	29	34.5%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	260	42.6%
Completed high school or higher education	1319	27.8%
p values (Pearson chi-square)		<0.0001

CSHCN status
Non-CSHCN 1464 29.5%
CSHCN 175 31.0%
p values (Pearson chi-square) - 0.47
At risk for developmental delay (online only)
Low/No risk 516 23.0%
High/Moderate risk 225 30.2%
p values (Pearson chi-square) - <0.0001

KPNW study: After controlling for other child and family demographic and health factors and provider characteristics, the likelihood (or adjusted odds ratio-AOR) that a child met quality measure criteria differed significantly according to child's age.

Table 1b.4d: Mean number of developmental services care components for which quality care was received and the proportion of children meeting criteria for receiving quality developmental services by characteristics of children and families.

Characteristic of Child or Child's Family % Meeting SAF Criteria
Child's Age
Less than 9 mos. 63.1% S
10 to 18 mos. 55.8%
19 to 49 mos. 47.4%
AOR: .55
Child's Gender
Male Child 52.6% NS
Female Child 53.7%
Child's Race
White, Non-Hispanic 52.0% BS
Asian, Non-Hispanic 62.0%
AOR: 1.47
Hispanic 53.2%
Other Race,
Multiple Race 56.7%
Birth Order
Not First Born 54.7% NS
First Born 52.0%
AOR: .80
Child's Risk for Developmental, Behavioral or Social Delays (Using Parent's Evaluation of Developmental Status)
Low/No Risk 52.8% NS
At Risk 54.0%
Respondent Education
More than High School 52.9% NS
High School or Less 54.9%
Respondent's Risk for Depression (Using the Kemper Screener)
No Symptoms of Depression 52.9% NS
Symptoms of Depression 56.9%
NOTE: Adjusted odds ratios (AOR) derived from regression analyses listed in the table are shown only if they are statistically significant. AOR uses the first subgroup of each characteristic as a reference. s differences significant at the $p < .05$ level of significance; BS differences significant at the $p < .10$ level; NS differences not significant.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4**
NA

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

De.7. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)
This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** CAHMI_Data_Dictionary_Assessment_ASA_Use_Safety.pdf

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.
No

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.
None

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator measures the number of parents who had a well child visit within the last 12 months and who were asked about alcohol use, substance abuse, safety and firearms in the house.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator is the number of survey respondents (parents) answering “Yes” to all three questions in the Assessment of Family Alcohol Use, Substance Abuse and Safety section of the Promoting Healthy Development Survey (PHDS). An aggregated 100% response of "Yes" is needed to achieve quality for this aspect of care.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The denominator is the number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months and answered all of the Family Alcohol Use, Substance Abuse and Safety questions on the Promoting Healthy Development Survey(PHDS, see Attachment A-2, page 17).

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The denominator is the number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months and who answered all of the Alcohol Use, Substance Abuse and Safety Assessment questions on the PHDS (see Attachment A-2, page 17).

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Missing data were excluded from the analysis.

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
Surveys in which two or more questions from the Family Alcohol Use, Substance Abuse and Safety section of the PHDS were missing were excluded from analysis. Approximately 2.6% of parents who started the Online PHDS did not complete the survey (range 0.0-5.6% for top 5 providers with highest number of surveys; see Testing form, pages 21-23 for more detailed information on missing data).

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*
Although no stratification is required, the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the findings by possible vulnerability, should any individual provide have sufficient data (parent responses) to do so. Potential variables for stratificationners include:
(1) Child demographic characteristics (e.g., the child´s age, race);
(2) Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or
(3) Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
No risk adjustment or risk stratification
If other:

S.12. **Type of score:**
Rate/proportion
If other:

S.13. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
The numerator is the sum of survey respondents (parents) answering "Yes" to all three of the Family Alcohol Use, Substance Abuse and safety questions. The denominator is the sum of all respondents answering all of the assessment questions. Surveys missing one or more responses to this set of questions are excluded from analysis. An aggregated score of 100% is required for achieving quality for this measure.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
NA

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Data are collected using the parent-reported "Promoting Healthy Development Survey" (PHDS) developed by the CAHMI (www.wellvisitsurvey.org). Instructions for survey completion are included with the survey. Family Centered Care questions are multiple choice (Yes Definitely, Yes, Somewhat, and No). The PHDS is survey is initiated by the provider who can send it to all parents who have received a well child visit. CAHMI has a website (www.phdstoolkit.org) where providers can register to use for the PHDS. This site assigns each provider a unique URL, which allows for provider identification by CAHMI as well as light branding with the provider´s logo so that it is identifiable by the parent. The PHDS Toolkit website sends an email to the provider with the unique URL link to the survey. The provider then sends the link to the parents asking them with instructions to fill out the survey and provide feedback about the visit. The parent fills out the survey and receives a customized feedback report. The survey data are captured on a secure HIPAA compliant CAHMI server. Through the PHDS Toolkit website, providers can generate a report that aggregate parent data information from the survey. Providers must have a minimum of 10 surveys to generate a report to maintain parent confidentiality. See Evidence Form, Figure 1 for a visual model this process.

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Other

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.
The Assessment of Family Alcohol Use, Substance Abuse and Safety measure is included as part the CAHMI Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org). The data are generated by parents filling out the PHDS. The PHDS is based in English. See Evidence Form, Figure 1 for a description visual model of the data collection process.

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Individual

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Clinician Office/Clinic
If other:

S.22. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
NA

2. Validity – See attached Measure Testing Submission Form
CAHMI_NQF_testing_attachment_Assesssment_Alcohol_Drug_Safety_020217.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)
Yes

2.2 For maintenance of endorsement

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)
No

2.3 <u>For maintenance of endorsement</u>
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care)

What were the statistical results of the analyses used to select risk factors?

Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)
No - This measure is not risk-adjusted

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2968
Measure Title: Assessment of Family Alcohol Use, Substance Abuse and Safety
Date of Submission: 2/2/2017
Type of Measure:

☒ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process	☐ Efficiency
☐ Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record

☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Patient reported data	☒ other: Patient reported data

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
We used the following data sources for testing of the quality measure:

1) Online Promoting Healthy Development Survey (PHDS) – data collected through an online, publicly available tool (Promoting Healthy Development Survey-PHDS). Parents who had a well-child care visit in the last 12 months can complete the PHDS. Providers initiate the survey. (See Evidence Form Figure 1 for visual model of the Online PHDS.)
2) Kaiser Permanente Northwest (KPNW) Study – CAHMI partnered with Kaiser Permanente Northwest in Portland, Oregon. The study aimed to evaluate the level and variations in the quality of preventive and developmental services for young children and assess the contribution of key system, provider and patient factors.
3) HRSA Evaluation Study - The specific goal of this study was to evaluate the feasibility, acceptability and impact of three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young children in the context of discussions between pediatric clinicians and parents during well-child visits. The evaluation measures used data from 5 different tools/surveys including PHDS. The parent-completed PHDS was administered before and after the intervention to assess changes in the quality of well-child care. The study funded by Health Resources and Services and Administration's (HRSA) Maternal and Child Health Bureau. (Patient Centered Quality Improvement of Well-Child Care, Final Report, Supported by a grant from the Maternal and Child Health Bureau Research Grants Program, Health Resources and Services Administration, R40 MC08959 03-00.)

1.3. What are the dates of the data used in testing? 2004-2016
Online PHDS: 2008-2016
KPNW Study: 2004-2005
HRSA Evaluation Study: 2010-2012

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Online PHDS: n=5,670 surveys reporting on quality of care provided by 299 individual pediatricians and primary care providers from 88 clinics in 36 states. Participation is a voluntary self-selection process based on knowledge and interest in quality improvement in their practice.

KPNW Study: Provider-level surveys and quality of care assessment were focused on the care provided by 56 individual providers (44 pediatricians, 9 nurse practitioners, 3 physician assistants) in the pediatrics department who were organized into ten geographically distinct offices.

HRSA Evaluation Study: Three pediatric offices in Oregon: 1) a rural site, (4 pediatricians), 2) an urban site (8 pediatricians), and 3) an urban site, (12 pediatricians). All pediatricians in selected clinic and office staff participated in relevant baseline and follow up data collection.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

<u>Online PHDS:</u> Children age 3-48 months of age whose parents completed the online publicly available PHDS were included in the testing. During 2008-2016, we received 5,670 completed surveys. Of those surveys, 5,355 surveys with provider IDs were used for analyses. Children's socio-demographic and health characteristics varied across the individual providers included in the analysis.

Table 1.6a: Characteristics of children for whose visited provider ID is available

Characteristics	All Children (n=5355)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
Age of child						
Under 10 months of age	38.3%	19.1%	49.5%	33.3%	54.4%	24.7%
10 to 18 months of age	34.7%	39.4%	38.5%	38.9%	29.1%	57.1%
19-47 months of age	27.0%	41.5%	12.1%	27.8%	16.5%	18.2%
Race/ethnicity of child						
White, non-Hispanic	53.8%	13.3%	81.0%	20.3%	50.7%	17.3%
Hispanic	40.8%	81.1%	14.3%	74.7%	40.8%	78.7%
Other race/ethnicity	5.3%	5.5%	4.7%	7.0%	8.4%	4.0%
Respondent education level						
Did not complete high school	12.1%	23.6%	0	34.1%	6.4%	15.8%
Completed high school	88.9%	76.4%	100%	65.9%	93.6%	84.2%
Children who qualify for Children with Special Health Care Needs (CSHCN) Screener criteria						
CSHCN	10.1%	7.4%	8.8%	10.0%	11.4%	5.2%
Non-CSHCN	89.9%	92.6%	91.2%	90.0%	88.6%	94.8%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	22.7%	-	24.4%	-	28.9%	0%

-Data is not available due to small sample size

KPNW Study: The population studied was children 3 to 48 months old who live in a metropolitan area in the Pacific Northwest. One randomly selected child per household whose age would be no younger than 3 months of age and no older than 48 months of age at the time that their parents received the survey and had one or more well-child visits were eligible to be sampled. A random sample of 5,755 children were identified. Of the 5,755 sampled children, 2,173 surveys were returned (37.8%).

Table 1.6b: Characteristics of children for whom survey responses were received, KPNW study, Top 5 individual providers with highest number of surveys

Characteristics	All Children (n=2173)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		7 (n=80)	53 (n=77)	4 (n=74)	1 (n=67)	43 (n=66)
Age of child						
Under 10 months of age	22.0%	20.0%	19.5%	20.3%	22.4%	21.2%
10 to 18 months of age	26.6%	25.0%	29.9%	35.1%	22.4%	15.2%
19-47 months of age	51.4%	55.0%	50.6%	44.6%	55.2%	63.6%
Gender of child						
Female child	46.2%	48.8%	49.4%	47.3%	41.8%	45.5%
Male child	53.8%	51.3%	50.6%	52.7%	58.2%	54.5%
Race/ethnicity of child						
White, non-Hispanic	72.9%	84.8%	77.0%	93.2%	76.9%	62.5%
Asian, non-Hispanic	7.8%	2.5%	6.8%	1.4%	3.1%	20.3%
Hispanic	8.9%	6.3%	12.2%	2.7%	10.8%	10.9%
Other race/ethnicity	10.4%	6.3%	4.1%	2.7%	9.2%	6.3%
Child is the first born in the family	52.1%	52.5%	40.8%	35.1%	54.5%	52.3%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	31.3%	21.5%	24.7%	27.0%	29.7%	26.2%
Education level of mother						
High school or less	12.7%	20.3%	3.9%	14.9%	16.7%	6.2%
More than high school	87.3%	79.7%	96.1%	85.1%	83.3%	93.8%

HRSA Evaluation Study: The study inclusion criteria were used to determine which parents/guardians of children were invited to participate in the interventions and/or evaluation from each participating study site:

- Parent has a well-child visit scheduled at this intervention site for one or more of their children.
- The child is scheduled for their 4-month to 3-year-old well-child visit and, therefore, is between the ages of 4 and 40 months (e.g. 40 month old children could be there for their 3 year well-child visit)
- The parent can read and understand English and is able to complete the intervention and evaluation tools.
- For intervention, the parent was able to access the online version of the Plan My Child's Well-Visit tool and the online evaluation survey.

The analysis includes 551 completed surveys at baseline (2010) and 275 completed surveys at follow-up (2011-12)

Table 1.6c. Sample description for baseline and follow-up PHDS respondents

Characteristics	Baseline (n=551)	Follow-up (n=275)
Visit type of child for whom survey was completed		
4, 6 or 9-month	38.9%	36.2%
12, 15 or 18-month	33.7%	41.3%
24 or 36-month	27.4%	22.4%
Birth order of child for whom survey was completed		
First child	42.2%	56.6%
Not first child	57.8%	43.4%
Race/ethnicity		
White, non-Hispanic	80.3%	83.5%
Hispanic	8.4%	6.6%
Other/multiple, non-Hispanic	8.6%	6.6%
Asian, non-Hispanic	2.7%	3.3%
Insurance type		
Private or private and public	90.7%	86.7%
Public only (includes Medicaid, Medicare, CHIP and Military)	7.6%	12.1%
Other	0.7%	0.4%
None	0.9%	0.8%

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Online PHDS and KPNW study data were used for reliability testing and stratification analysis. Validity findings are presented from a peer-reviewed publications and online PHDS data. Performance analysis was conducted using the online PHDS, KPNW study and HRSA Evaluation Study data.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Online PHDS: Child's age, sex, race/ethnicity, and respondent (parent) age, race/ethnicity, and education level. The survey does not have a question asks about family income due to complexity of collecting income data by self-reported survey. However, the online PHDS has items assessing the family's economic situation: How much trouble does the family have paying for a) child's health and medical expenses; b) supplies like formula, food, diapers, clothes and shoes; and c) health care for the parent.
KPNW Study: Child's age, sex, race/ethnicity, and education level of mother
HRSA Study: Child's age, race-ethnicity, and insurance type

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Internal consistency: Cronbach's alpha was calculated using the Online PHDS and KPNW data. Cronbach's alpha is the most widely used in health care research when multiple-item measures of a concept or construct are employed. The acceptable values of alpha ranges from 0.70 to 0.95.

The primary aim of the quality measure is to detect difference between providers on the quality of care provided to young children. Provider level reliability was assessed by inter-unit reliability (IUR) using analysis of variance. IUR can be interpreted as the fraction of the variation among provider scores that is due to real differences, rather than due to chance. If the IUR is higher, the ability of the item or scale measure to discriminate across programs is greater. Scales with reliability coefficients above 0.70 provide adequate precision for use in statistical analysis of unit-level comparisons.[1] As the IUR gets smaller, a larger sample is needed in order to reliably discriminate across programs. In the analysis we included providers with 10 or more completed surveys.

Intra-class correlation (ICC) was calculated using ANOVA, as a ratio of the variance between groups over the total variance. The interpretation of the ICC is as the proportion of relevant variance that is associated with differences among measured objects.[2] Fleiss (1981) and Cicchetti and Sparrow (1981) from the medical group state that ICC range categories are: < 0.40 = poor; 0.40 – 0.59 = fair; 0.60 – 0.74 = good; and > 0.74 = Excellent[3]. Values above about 0.7-0.8 are considered acceptable for applied tests. In the analysis we included providers with 10 or more completed surveys.

1. Nunnally, J. C. Psychometric theory (2nd ed). 1978, New York: McGraw-Hill.
2. McGraw, K. O., & Wong, S. P. Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1996:1(1), 30-46.
3. Cicchetti D.V., and Sparrow, S.S. Developing criteria for establishing the interrater reliability of specific items in a given inventory. American Journal of Mental Deficiency, 1981:86, 127-137.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 2a2.3a. Assessment of Family Alcohol Use and Substance Abuse and Safety (AFAUSAS): Content, Scoring and Internal Consistency, Online PHDS, all providers and top 5 individual providers with highest number of surveys

What is measured	Scoring	Internal Consistency (Cronbach's Alpha)					
This composite measure assesses the degree to which providers ask about (1) feeling safe in the home; (2) alcohol use and substance abuse use; and (3) firearms in the home.	Mean score on a multi-item scale	All providers	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
			1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
		0.81*	0.82*	0.80*	0.79*	0.80*	0.69

*Met criteria for reliability and internal consistency.

The Chronbach's alpha for the Assessment of Family Alcohol Use and Substance Abuse and Safety measure is 0.81, ranging 0.69-0.82 across the providers with highest number of surveys. These findings are consistent with the findings of the previous peer-reviewed publications.[4,5] Inter-unit reliability coefficient for the measure scale is within the recommended threshold (0.71) to reliably detect difference between providers. Intraclass correlation coefficient for the measure is 0.78, indicating that 77.7% of the variance in the mean of the providers is "true" rather than due to chance.

4. Bethell C, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. Pediatrics. 2001 May;107(5):1084-94.
5. Christina Bethell, PhD, MPH, MBA; Colleen H. Peck Reuland, MS; Neal Halfon, MD, MPH; Edward L. Schor, Measuring the Quality of Preventive and Developmental Services for Young Children: National Estimates and Patterns of Clinicians' Performance. Pediatrics, 2004, 113(6):1973-83

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Psychometric analyses demonstrated that the Assessment of Family Alcohol Use and Substance Abuse and Safety quality measure scale have strong internal consistency (Cronbach's alpha ranged 0.69-0.82 across individual providers) and reliability detect differences between providers (IUR coefficient 0.71 and ICC 0.78). The quality measure score provides psychometrically reliable assessment of the provision of nationally recommended well-child care with strong internal consistency.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

A standard, multistage process was used to ensure validity of the PHDS items/measures:

- Focus groups and in-depth cognitive interviews were conducted throughout the survey development process;
- A review of literature identified through Medline or during key informant interviews; and,
- Three Advisory Groups comprised of pediatricians, family practitioners, consumer representatives, public health experts, and researchers, regularly reviewed and provided input on the identification of quality measurement topics and the development of the PHDS.

A "gold standard" does not exist for determining the criterion validity of patient-reported measures of quality. However, to ensure the validity of the PHDS quality measure results, we followed rigorous procedures representing best practices within the field to develop the survey questions. To ensure the content validity of measures of parent experiences, we used qualitative methods, including both focus groups and cognitive interviews, to inform development and evaluation of the AFAUSAS questions.

Focus groups with families aimed to identify the aspects of health care quality that are important to parents in the area of preventive care for their children. In-depth cognitive testing of the draft survey items was conducted with 15 families representing a range of racial, income and education groups as well as different types of health insurance coverage, age of child, age and sex of parent, and number of children in family. Focus groups and cognitive interviews with 35 health care providers in Vermont and Washington and 20 parents of young children in Vermont were conducted to inform item-reduction, administration specifications, and reporting templates. Survey modifications were made based on findings in order to improve the reliability, validity and cognitive ease of the AFAUSAS items.

Factor analysis was conducted to assess the construct validity of the quality measure. A Scree test was used to determine the number of factors to extract. Both oblique and orthogonal rotations were evaluated with promax and varimax methods used, respectively.[1] Acceptable level of factor loading for instruments developed for research purposes can be as low as 0.60[2] and factor loading more than this threshhold is considered as a strong association.[3]

To assess the concurrent validity of the quality measure, hypothesized associations among PHDS items and scales were examined. We tested a hypothesis that respondents who indicate that providers talked with them about keeping house and car safe topics more likely to report increased confidence as a parent because of interactions with health care providers compared with respondents who indicate that providers did not talk with them.

Pearson correlation coefficients were calculated across all scale measures to test hypotheses about expected relationships among the PHDS quality measures and to assess the degree to which each of the PHDS quality measures provide unique information. We expect a moderate or strong correlation between the family assessment scale measures (>0.30) and inter-scale correlation coefficients to be less than 0.80.

1. Bethell C, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. Pediatrics. 2001 May;107(5):1084-94.

2. Suhr D and Shay M. Guidelines for reliability, confirmatory and exploratory factor analysis. Accessed at: http://www.wuss.org/proceedings09/09WUSSProceedings/papers/anl/ANL-SuhrShay.pdf. Retrieved 02/01/2017
3. Costello A.B and Osborne J.W. Best Practices in Exploratory Factor Analysis: Four recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation. 2005:10(7). Accessed at: http://www.pareonline.net/pdf/v10n7.pdf, Retrieved 02/01/2017

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Using behavior coding methods, for each item in the AFAUSAS quality measure, instances where the respondent required clarification or did not appropriately answer an item were noted. Also, items where the interviewer had difficulty asking the question without edits to the wording were noted. Data analysis was used to inform item-reduction. Content was revised and refined iteratively with each set of interviews.

Cognitive testing confirmed the readability of the AFAUSAS items for people across a range of educational levels. Parents were uniformly able to complete the self-administered survey in 10-15 minutes. Readability assessments indicated the items to be written at the 8th-9th grade reading level. Survey design and formatting was finalized with input from a group of experts and family representatives.

Factor analysis demonstrated a strong factor structure within the AFAUSAS measure. Each of the items used to construct the AFAUSAS quality measure was used in the factor analysis. Average factor loading for AFAUSAS was 0.66. Inter-item correlation ranged between 0.54-0.60. Factor analysis suggests that the scale items are unidimensional.

Concurrent validity testing showed that parent more likely to report “a lot or a little” more confident in protecting child from injury if provider talked with parent about keeping house and car safe (odds ratio [OR]: 5.9, 95% confidence interval [CI]: 3.4-10.2; OR: 8.3, 95% CI: 5-13.8). In addition, parents report that they are rarely asked about other psychosocial issues, including gun safety or how parenting works into their daily activities. These findings are similar to those of other studies.[4,5,6]

Correlations between the PHDS quality measures were not so high as to suggest redundancy across measures (average correlation: 0.34). As expected, the highest correlation observed was between the “Assessment of family psychosocial well-being” & “Assessment of smoking, drug and alcohol use and safety in the family” (0.54) and “anticipatory guidance from providers” & the “family-centered care” measures (0.52).

Table 2b2.3. Pearson Correlation Coefficients among PHDS Quality Measures (online PHDS)

Scale Measures	Anticipatory Guidance and Parent Education	Family Centered Care	Ask About Parental Concern	Assessment of smoking, drug and alcohol use and safety in the family	Assessment of family psychosocial well-being
Family Centered Care	.52				
Ask About Parental Concern	.16	.14			
Assessment of smoking, drug and alcohol use and safety in the family	.16	.13	.07		
Assessment of family psychosocial well-being	.19	.16	.09	.54	

Average correlation: 0.34

Some AFAUSAS items have been used in the National Survey of Children's Health. he PHDS-derived quality measures are among the few recognized in the Agency for Healthcare Research and Quality's Child Health Toolbox and the National Quality Measures Clearinghouse as measures that meet basic criteria for use as standardized indicators of health care quality for children.

4. Young KT, David K, Schoen C. Listening to parents. A national survey of parents with young children. Arch Pediatr Adolesc Med. 1998;152: 255–262
5. Kahn RS, Wise PH, Finkelstein JA, et al. The scope of unmet maternal health needs in pediatric settings. Pediatrics. 1999;103:576 –581
6. Kemper KJ, Osborn LM, Hansen DF, Pascoe JM. Family Psychosocial screening: Should we focus on high-risk settings? J Dev Behav Pediatr. 1994;15:336 –341

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The AFAUSAS quality measure provides conceptually and psychometrically valid assessment of the provision of nationally recommended preventive care services for young children, with strong construct validity (average factor loading: 0.66). Each of the PHDS quality measure provides unique information about performance.The measure is used in national surveys and recognized as a measure that meet basic criteria for use as standardized indicators of health care quality for children. The measure serves as an important complement to existing quality measures.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Not applicable

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Not applicable

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Not applicable

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☒ **Stratification by** variable number of **risk categories**

☐ **Other,**

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

Not applicable

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

This measure does not require risk adjustment because we do not expect variation in the quality of care provided for children due to risk factors, e.g. children with special health care needs. The provider's performance should be the same regardless of risk factors. The national experts extensively reviewed the risk adjustment requirements during development of the measure items of the PHDS tool and did not recommend risk-adjustment for the measures. In addition, during the KPNW study, we did assessment of whether the probability of receiving guidance, education or screening was higher according to a child's level of need or risk, thereby indicating that providers are customizing care to children. The study found no evidence emerged that providers customize care to children most at risk.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Identification of variation in quality measures across subgroups of children helps to highlight aspects of care and population of children for which preventive and developmental services may be most need of improvement. Although no stratification is required (number of surveys for each individual providers may not be sufficient to stratify), the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the quality measures by possible vulnerability:

- Child demographic characteristics (e.g., the child's age, race)
- Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs)
- Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

Based on extensive literature review and expert panel, we identified that child and parent demographics such as age, sex, race-ethnicity, income, insurance, parent behavior, CSHCN screener and follow-up for children at risk can be used for stratification. Several studies have documented differences in access and quality of care provided to children, as well as in parent-reported satisfaction with care.[1-2] One study found: Overall, 94.0% of parents reported 1 or more unmet needs for a number of aspects of care, including assessing family alcohol use, substance abuse and safety. Uninsured

children and children aged 18 to 35 months are disproportionately represented among the 15.3% of children whose parents indicated an unmet need this area of care. There are significant variations in performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors.[3]

The KPNW study assessed child and family characteristics to characterize the child and their family based on the PHDS item responses: child's race/ethnicity, birth order, risk for developmental, behavioral, or social delays using responses to Frances Glascoe's Parents' Evaluation of Developmental Status (PEDS) items included in the ProPHDS 29 parent's education; and whether he/she is experiencing symptoms of depression using Kathy Kemper's screening items. Adjusted odds ratios were calculated using logistic regression analysis in order to assess differences in the odds of meeting quality measure criteria according to child, family and provider characteristics, after controlling for other variables.

References:

1. Halfon N, Regalado M, Sareen H, Inkelas M, Reuland CH, Glascoe FP, Olson LM. Assessing development in the pediatric office. Pediatrics. 2004 Jun;113(6 Suppl):1926-33.
2. Weech-Maldonado R, Morales LS, Spritzer K, Elliott M, Hays RD. Racial and ethnic differences in parents' assessments of pediatric care in Medicaid managed care. Health Serv Res. 2001 Jul;36(3):575-94.
3. Bethell C, Reuland CH, Halfon N, Schor EL. Measuring the quality of preventive and developmental services for young children: national estimates and patterns of clinicians' performance. Pediatrics. 2004 Jun;113(6 Suppl):1973-83.

2b4.4a. What were the statistical results of the analyses used to select risk factors?
Not applicable

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)
See 2b4.3.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)
Pearson's chi-square test was used to compare the prevalence of AFAUSAS quality measure across the stratification characteristics. We preformed logistic regression analysis in order to assess differences in the odds of meeting quality measure criteria according to child, family and provider characteristics, after controlling for other variables.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Not applicable

2b4.9. Results of Risk Stratification Analysis:

Online PHDS: Variation is observed according to a child's age; race/ethnicity; level of risk for developmental, behavioral, or social delays. Non-Hispanic white children are less likely to meet criteria on the Family Assessment measures. Children of lower educated mothers and children at high risk for developmental delay are more likely to have high Family Assessment scores.

Table 2b4.9a. Assessment of smoking, drug and alcohol use and safety (asked about all items) in the family by child demographics and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	705	31.3%
9-18 months	531	28.1%
19-48 months	403	29.1%
p values (Pearson chi-square)	-	0.06
Gender		
Male	215	34.7%
Female	210	33.6%
p values (Pearson chi-square)	-	0.67
Race/ethnicity		
Hispanic	723	37.1%
White non-Hispanic	680	23.9%
Black non-Hispanic	32	32.0%
Asian non-Hispanic	29	26.1%
Other/Multi race, non-Hispanic	29	34.5%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	260	42.6%
Completed high school or higher education	1319	27.8%
p values (Pearson chi-square)		<0.0001
CSHCN status		
Non-CSHCN	1464	29.5%
CSHCN	175	31.0%
p values (Pearson chi-square)	-	0.47
At risk for developmental delay (online only)		
Low/No risk	516	23.0%
High/Moderate risk	225	30.2%
p values (Pearson chi-square)	-	<0.0001

KPNW study: After controlling for other child and family demographic and health factors and provider characteristics, the likelihood (or adjusted odds ratio-AOR) that a child met quality measure criteria differed significantly according to child's age.

Table 2b4.9b: Mean number of developmental services care components for which quality care for family assessment of alcohol use, substance abuse and safety was received and the proportion of children meeting criteria for receiving quality developmental services by characteristics of children and families.

Characteristic of Child or Child's Family	% Meeting SAF Criteria
Child's Age	
Less than 9 mos.	63.1% [S]
10 to 18 mos.	55.8%
19 to 49 mos.	47.4% AOR: .55
Child's Gender	
Male Child	52.6% [NS]
Female Child	53.7%
Child's Race	
White, Non-Hispanic	52.0% [BS]
Asian, Non-Hispanic	62.0% AOR: 1.47
Hispanic	53.2%
Other Race, Multiple Race	56.7%
Birth Order	
Not First Born	54.7% [NS]
First Born	52.0% AOR: .80
Child's Risk for Developmental, Behavioral or Social Delays (Using Parent's Evaluation of Developmental Status)	
Low/No Risk	52.8% [NS]
At Risk	54.0%
Respondent Education	
More than High School	52.9% [NS]
High School or Less	54.9%
Respondent's Risk for Depression (Using the Kemper Screener)	
No Symptoms of Depression	52.9% [NS]
Symptoms of Depression	56.9%

NOTE: Adjusted odds ratios (AOR) derived from regression analyses listed in the table are shown only if they are statistically significant. AOR uses the first subgroup of each characteristic as a reference. [S] differences significant at the $p < .05$ level of significance; [BS] differences significant at the $p < .10$ level; [NS] differences not significant.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

The demographic and socioeconomic survey items included in the quality measure make it possible for providers to identify populations and subgroups for which health service delivery improvement is most needed.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)
Not applicable

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Differences in the quality measure scores across providers is demonstrated for (1) 5 top individual providers with the highest number of surveys using Online PHDS data; (2) across 56 providers using KPNW study data; and (3) pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.

Online PHDS: The performance scale for the quality measure was calculated using the scoring methods described in Attachment A-4. Individual provider level differences in performance were illustrated by the proportion of children meeting the quality of care criteria across the 5 providers with the highest number of completed surveys after their well-child visit.

KPNW Study: The significance of differences observed in the proportion of children meeting criteria for the quality measure across pediatric providers (n=56) was evaluated using t-tests. The relative spread in the quality measure score across providers was assessed using the coefficient of variation statistics (standard deviation across providers multiplied by 100%). Multi-level regression models were conducted using the pediatric provider as the level 2 clustering variable, in order to assess the degree to which the probability that a child meets criteria on each quality measure is explained by differences between providers (called the "clustering effect"). In implementing this multi-level regression method (Empty Model), the presence of a significant clustering effect by pediatric providers was estimated prior to accounting for the child and family characteristics associated with each provider. Second, variables related to the child and family characteristics (child's age, gender, race/ethnicity, birth order, developmental and behavioral delay risk status; parent education and risk for depression) were added to the Empty Model to assess how much of the provider clustering effect observed remains after accounting for these characteristics (called the "Patient Model").

HRSA study: Quantitative data results for the baseline (2010) and follow-up (2011-12) study of the intervention sites using the HRSA Evaluation Study data were conducted using basic descriptive statistics to describe each sample and applying chi-square test of statistical significance to assess differences in the quality measure for the baseline and follow-up samples.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)
Online PHDS: Table 4a present the proportion of children whose care met for the quality measure of family assessment for alcohol use, substance abuse and safety across the 5 providers with the highest number of surveys. Variation across providers who asked parents about all survey items related to substance abuse and firearms in the home is substantially wide across observed providers. Only 7.7%-48.2% of parents of young children reported that their child's pediatric clinician discussed psychosocial topics including alcohol use, substance abuse and safety.

Table 2b5.2a: Proportion of Children Meeting Measure Criteria, Top 5 individual providers with highest number of surveys

Characteristics	All Children (n=5355)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
Asking parents about substance abuse and firearms in the home (at least one of these topics)	60.5%	59.8%	24.2%	82.4%	70.5%	77.0%
Asking parents about substance abuse and firearms in the home (asked all items)	29.6%	29.3%	7.7%	48.2%	38.5%	31.1%

KPNW Study: 53.1% of children had parents reporting that providers discussed at least one alcohol or substance abuse and safety topic. Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant (p=0.04)

Table 2b5.2b: Proportion of all children in the study who met criteria for receiving quality services and ranges in proportion across providers. (SD=Standard Deviation)

Developmental Services Quality Measures	Proportion of All Children Meeting Measure Criteria (n = 2173)	Range in the Proportion of Children Meeting Measure Criteria Across 51 Pediatric Providers	Relative Variation (COV) in Measure Scores Across Pediatric Providers
Asking parents about substance abuse and firearms in the home (at least one of these topics)	53.1%	32.4% to 69.8% SD: 9%; (p = 0.04)	17.9%

Only providers with n=15 or more PHDS responses are included in the provider level analysis. Provider level n ranges from 15 to 153.

Multi-level analysis: For the Empty Model that used the provider as the level 2 clustering variable, only 1.1% to 2.2% of the total variance observed in whether children met criteria for each of the all quality measures was explained by either measured or unmeasured differences between the providers that they see. This suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers. These findings translate into a 1.19 to 1.29 median odds ratio across the six quality measures, indicating that the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across the all quality measures. When child/family level characteristics are added to the model (Patient Model), the total variance explained by differences between providers does not change significantly.

HRSA study

The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to be asked about one or more psychosocial (family assessment) topics, including alcohol use, substance abuse and safety issues, at follow-up. The tables below present comparison of percent of children who received care met the quality care criteria between baseline and follow-up survey data.

Table 2b5.2c. Family Assessment*, by Children's Characteristics

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	23.9% (132)	45.0% (145)	<0.0001
10-18 months	21.5% (87)	34.4% (111)	<0.0001
19-48 months	29.3% (103)	50.5% (112)	<0.0001
Race			
Hispanic	26.0% (26)	47.8% (22)	0.01

White	24.5% (252)	42.7% (305)	<0.0001
Asian	28.6% (8)	(4)	-
Multiple or other	50.0% (12)	50.0% (9)	1.00
Insurance type			
Private or private and public	22.9% (248)	37.9% (257)	<0.0001
Public only (includes Medicaid, Medicare, CHIP, and Military)	33.5% (64)	39.1% (63)	<0.0001
Other insurance type	(2)	(1)	N/A
Uninsured	(4)	(3)	N/A
At risk of developmental delay			
Low/no risk	24.3% (248)	41.6% (227)	<0.0001
High/moderate risk	26.0% (73)	44.7% (76)	<0.0001

* Parent was asked about one or more family assessment topics

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

Significant gaps and unexplained variations remain in the quality of family assessment services for families with young children. The probability of receiving quality care varies nearly as much across children seeing the same provider as across providers. The quality measure assessed here provide a relatively comprehensive picture of performance in the area of provider assessment of family alcohol use, substance abuse and safety.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)
Not applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Table 2: Proportion of all children in the study who met criteria for receiving quality developmental services across six components of care and ranges in proportion across providers and offices. (SD=Standard Deviation)
Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
The quality measure items were developed based on several rounds of cognitive interviews with parents to ensure quality of responses appropriate to the questions and minimize missing responses.

Online PHDS: Rate of survey completion was calculated based on survey start and complete dates for each respondent. According to the quality measure scoring protocol, if a parent answered less than half of the items in the AFAUSAS measure, their score is considered to be missing. This does not include items that should have been appropriately skipped. Missing responses are not given a valid score and are not included in the calculation of the quality measure.

KPNW Study: Of the 5,755 sampled children, 2,173 surveys were returned (37.8%). For these children, the provider the parent identified and the provider to which the child was assigned by the health plan were the same 97.3% of the time. A 95% response rate was obtained for the provider survey.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Online PHDS data show that overall 2.6% of parents who started the survey did not complete the survey. Table 2b7.2a presents the frequency of missing values for the Assessment of Family Alcohol Use, Substance Abuse and Safety measure.

Table 2b7.2a. The frequency of missing values for Assessment of Family Alcohol Use, Substance Abuse and Safety measure, overall and top 5 providers

Quality measures	Overall	Provider ID				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Assessment of Family Alcohol Use, Substance Abuse and Safety	2.5% (142)	2.1 (2)	0	5.6 (5)	1.3 (1)	3.9 (3)

KPNW study: Children whose parents responded were not different from those who did not respond in terms of their gender and insurance type. The responding population were slightly less likely to be in the 19 to 48 month age group (55.8% sampled, 51.5% responding) and were somewhat more likely to have had more than one well-visit in the past (67.5% sampled, 74.7% responding).

Table 2b7.2b. Sociodemographic Characteristics of KPNW Starting and Responding Sample

Characteristic	Proportion of Starting Sample (N=5755)	Proportion Respondents as of (N=2162)
Gender of Child[NS]		
Male child	52.7	53.7
Female child	47.3	46.3
Age of the Child[S]		
Child age 3-9 months	19.4	21.8
Child age 10-18 months	24.9	26.7
Child age 19-48 months	55.8	51.5
Type of Insurance[NS]		
Private	98.6	98.5
Public	1.4	1.5
Child's Health care utilization		
Number of well-child visits [S]		
1 Well-Child Visit	32.5	25.3
2 or More Well-Child Visits	67.5	74.7
Number of emergency room/urgent care visits		
0 ER/urgent care visits	49.8	51.0
1 ER/urgent care visit	26.2	25.8
2 or more ER/urgent care visit	24.0	23.2
Number of overnight hospital stays [NS]		
0 overnight hospital stays	96.6	96.9
1 or more overnight hospital stays	3.4	3.1

[S]Denotes variables for which statistically significant variation exists between the starting and responding sample for the target child or respondent characteristic.
[NS]No significant variation exists between the starting and responding sample for the target child or respondent characteristic.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Information about non-respondent is not available to compare with those who responded the survey because online PHDS is publicly available tool. However, the low rate of incomplete survey (2.6%) suggests that the measure was acceptable to respondents. Overall, the quality measure had 2.5% of missing cases, ranging 0-5.6% across the top 5 providers with the highest number of surveys. Few overall missing values suggest that the measure level results unlikely to be biased by non-response to the survey questions.

.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**
Other
If other: Data are provided directly by parents through the CAHMI developed Promoting Healthy Development Survey. (PHDS)

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.
Patient/family reported information (may be electronic or paper)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).
Data are parent-report using the CAHMI developed Promoting Healthy Development Survey (PHDS). CAHMI captures the data at the provider level through a process described above and in the Evidence Form, Figure 1.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**
Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**
IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.
During 2012-2016 we have experienced some operational delays. in the past several years. In 2012, the provider feedback reports were not automated. When providers wanted a summary report, CAHMI had to manually create them. This was excessively time consuming and CAHMI did not have resources to continue the manual generation of the reports. We sought and received funding to automate the reports. Some difficulties with contractors and staff change-over caused major delays in the project. Then, CAHMI moved from the Oregon Health & Sciences University to Johns Hopkins University School of Public Health in 2014, and it was necessary to upgrade the CAHMI servers. No technical support was available for the transition which caused further delays. Additionally, the PHDS was originally developed in 2001; thus much of the coding and back-end technology for this tool was antiquated and ceased to function after the move. Consequently, and as a result of new improved technology, we have had to redesign the two PHDS related websites - the PHDS toolkit and the parent survey -- as well as the CAHMI PHDS database. Lack of funding caused delays. However, we anticipate launching the new PHDS in February 2017.

Time and cost of data collection are low: provider registration takes about 10 minutes and the parent survey takes about 15-20 minutes to complete. To date, implementation has been limited by lack of funding and resources for outreach, communication

and technical support. Our experience in the development and evaluation of the PHDS demonstrated a clear and compelling need to work closely with providers to overcome the many myths that both parents and providers have about patient-engagement quality improvement tools. For the PHDS to be adopted by providers, it is essential to demonstrate, for example, that tool adds value for both the parent and provider, that it fits into and typically improves work flow in the office; improves parent-provider communication, and most important, improve the quality and delivery of nationally recommended services for children. This can only really be accomplished by collaboration and partnership with providers.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**
NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Professional Certification or Recognition Program Quality Improvement (Internal to the specific organization)	

4a.1. **For each CURRENT use, checked above (update for maintenance of endorsement), provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

NA

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
The PHDS toolkit (www.phstoolkit.org) and the parent-reported PHDS (www.wellvisitsurvey.org) were used by 68 uniquely identified providers across the country through 2013. We are happy to provide a list of these providers to NQF if desired. In 2014, CAHMI moved from the Oregon Health & Sciences University, Portland OR to the Johns Hopkins University, Baltimore, MD. As a result of the move, and because both server and database technologies had rapidly evolved and improved over the past few years, it was necessary to upgrade our servers, which in turn caused some technical issues with the links between the provider toolkit, the PHDS, and the CAHMI PHDS database. Additionally, the PHDS was originally used to compare providers within a practice as well as between practices within a health system. The anticipated use of the Online PHDS is intended to provide feedback only for individual providers and at the clinic or practice level but not between providers. The combination of these factors led to a decision to upgrade and redesign the PHDS toolkit, PHDS database and Parent Survey. (The PHDS parent survey itself, however, remains fully operational, although use has been nominal from 2014-present, and can be accessed at www.wellvisitsurvey.org.) The redesign required additional time, IT and CAHMI staff resources and delays were incurred during

2014-2015. However, we are now in the process of finalizing the PHDS Toolkit and database redesign, which is anticipated to be completed and launched in February 2017.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)
The redesigned PHDS system (registration toolkit, parent survey tool and PHDS database) is anticipated to be completed and fully functional by February 2017. We have a communication and outreach plan to promote the PHDS as part of the CAHMI Cycle of Engagement (see PHDS_Final_Appendix, Item #24), which includes the CAHMI Well Visit Planner (www.wellvisitplanner.org) -- a free parent engagement tool that helps prepare parents for the upcoming well child visit – and the post-visit PHDS which assesses whether the parent received services in alignment with national guidelines as well as family centered care. We have been promoting the Cycle of Engagement in national meetings (AMCHP, PAS, APHA, AcademyHealth ARM, National Child Heath Policy Meeting, and more) over the past several years. We presented the Cycle of Engagement at the CMS Quality Meeting December 13, 2016 and have further plans to unveil the redesigned version at meetings in 2017. The WVP and PHDS have also been endorsed tools that meet requirements for Bright Futures implementation.

We have received substantial interest in the CAHMI parent-engagement tools (both the WVP and the PHDS) from and are in extensive conversations with a number of organizations and agencies including health systems, payers, provider organizations – (CMS/Medicaid, Title V, Head Start, Kaiser Permanente and others); professional associations such as the American Academy of Pediatrics, Bright Futures, National Medicaid Medical Directors, the Children's Hospital Association (CHA), AcademyHealth, Association of Maternal and Child Health Programs (AMCHP), CityMatCH, National Initiative for Children's Healthcare Quality (NICHQ), Autism Speaks, Prevent Child Abuse America; National Prevention Information Network (NIPN); national community-based programs and organizations; philanthropic funders; software platform and electronic medical records systems developers and family organizations. We are in the process of securing funding for Cycle of Engagement EMR integration and implementation projects in partnership with or from a number of interested parties. Further, we are finalizing our application to the American Board of Pediatrics to have the Online PHDS certified as a web-based Maintenance of Certification (MOC) (Part 4) quality improvement (QI) tool for pediatricians. ABP has expressed significant interest in the PHDS and provided some initial funding for the redesign efforts

Improvement
Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**
If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.
Based on PHDS feedback results from an evaluation study conducted in 2011-2012 in Oregon, Family assessment for one or more family assessment topics increased 103.3%, a statistically significant increase, from 21.5% at baseline (2010, n=116) to 43.7% post-assessment (2011-12, n=111, AOR: 3.32, 95% CI: 2.24-4.91).

4c. Unintended Consequences
The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**
There were no unintended or unexpected consequences that we are aware of.

4c.2. **Please explain any unexpected benefits from implementation of this measure.**
There were no unexpected benefits that we are aware of.

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
Extensive qualitative interviews with providers and parents have been conducted and previously reported (See Attachment 2, Evidence Report)

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
Key informant interviews and focus groups with parents and providers were held throughout the testing and evaluation period. We obtained baseline and post-implementation information from providers and post-implementation information from parents. It was necessary to work closely with practices to demonstrate value of the family engagement tools (Well Visit Planner and PHDS) as well as to modify the process to fit individual practice office culture and work flow. A significant amount of provider and staff education was needed to overcome fears and myths that the tool would add to, not help, time management and that parents would not want to participate. This was accomplished by continued and persistent relationship building, spending much time in the office setting with the staff and providers and holding frequent Q&A sessions as the process unfolded.

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.
Through key informant interviews and focus groups with parents and providers.

4d2.2. **Summarize the feedback obtained from those being measured.**
The Assessment of Family Alcohol Use, Substance Abuse and Safety measure is seen by providers as an excellent way by which they can improve the quality of the well child visit. In particular this matters a great deal to the providers who are being financially incentivized for family-centered care outcomes.

4d2.3. **Summarize the feedback obtained from other users**
For the most part, parents appreciated being asked about their experience with their well child visits and used it as a way to provide confidential feedback to the providers.

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
The feedback was helpful for future implementation efforts of CAHMI's family engagement tools. The feedback, however, did not result in any changes to the measure itself.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization of Related Measures
The measure specifications are harmonized with related measures;
OR

The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**
Are the measure specifications harmonized to the extent possible?
Yes

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**
NA

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)
NQF #0011 - the PHDS (Promoting Healthy Development Survey) - was endorsed by NQF on October 4, 2012. The PHDS contains the AFAUSAS measure. Neither the questions nor the scoring of the questions have changed since the PHDS was endorsed. It is not actually a competing measure; rather, the AFAUSAS measure is embedded in the PHDS tool.

Please note: The PHDS endorsement (#0011) can be found on the NQF measures website but does not appear to be found in the NQF directory in Question 5 above. Hence, we were forced to enter a "no" to Q5 in order to submit this application.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.
Attachment **Attachment:** Attachment_A_Supplemental_Materials_Revised_01_18_17-636203531160271953.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Child and Adolescent Health Measurement Initiative
Co.2 **Point of Contact:** Christina, Bethell, cbethhel@cahmi.edu, 443-287-5092-
Co.3 **Measure Developer if different from Measure Steward:** Child and Adolescent Health Measurement Initiative
Co.4 **Point of Contact:** Christina, Bethell, cbethhel@cahmi.edu, 443-287-5092-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**
Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.
National Advisors for Patient Centered Quality Improvement of Well-Child Care:
Betsy Anderson, Family Voices
David Bergman, Stanford University
Dimitri Christakis, University of Washington
Paula Duncan, University of Vermont
Cynthia Minkovitz, Johns Hopkins School of Public Health
Amy Perritti, American Academy of Pediatrics
Ed Schor, The Commonwealth Fund
Judy Shaw, University of Vermont
Sara Slovin, Johns Hopkins Medicine

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2002

Ad.3 **Month and Year of most recent revision:** 01, 2017

Ad.4 **What is your frequency for review/update of this measure?** 3 years

Ad.5 **When is the next scheduled review/update for this measure?** 01, 2018

Ad.6 **Copyright statement:** None

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Brief Measure Information

NQF #: 3223
Corresponding Measures:
Measure Title: Assessment of Family Psychosocial Screening
Measure Steward: Child and Adolescent Health Measurement Initiative
Brief Description of Measure: This measure is used to assess the proportion of children whose parents were assessed by a health provider on one or more of the recommended psychosocial well-being topics, including depression, emotional support, changes or stressors in the home, and how parenting is working.
Developer Rationale: Recommended developmental services, as set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau, include family psychosocial assessment and follow-up, which consists of maternal depression; mental health of parents; smoking, alcohol and drug use; presences of adequate economic, social, and emotional supports; guns; family violence; and other safety issues. In order to gauge the quality of recommended care provided, this type of information must be collected from the parent in order to identify the level at which providers discuss these issues with parents. Previous studies have shown that parents are willing to discuss such sensitive topics with providers.

Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child´s health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews which are not accurate for the specific level of information obtained in the Promoting Healthy Development Survey (PHDS).

Numerator Statement: The numerator is the number of parents who had a well child visit within the last 12 months and who were asked about psychosocial well-being.
Denominator Statement: The number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months and all answered questions related to the family psychosocial screening scale (see Attachment A-2, page 18).
Denominator Exclusions: Missing data are excluded from the analysis.

Measure Type: Outcome: PRO
Data Source: Other
Level of Analysis: Clinician : Individual

New Measure -- Preliminary Analysis

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcomes measure include providing rationale that supports the relationship of the health outcome to processes or structures of care. The guidance for evaluating the clinical evidence

asks if the relationship between the measured health outcome and at least one clinical action is identified and supported by the stated rationale.

Evidence Summary

- This is a Patient-Reported Outcome-Based Performance Measure (PRO-PM) derived from the responses to three questions on the Promoting Healthy Development Survey (complete survey starts on page 20 of the Appendix).
- The developer provided a logic model in both graphic and narrative: (1) the parent and child attend a well child visit with their provider; (2) the provider subsequently sends a survey -- the Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org), which includes one question (3 items, see Attachment A-2, page18) for the parent to complete; (3) when at least ten surveys have been completed, the provider receives a feedback report on parents' experiences of the visit and the extent to which they received an assessment of family psychosocial screening via the CAHMI PHDS Toolkit website (www.phdstoolkit.org); (4) the provider reviews the report and then can engage in a Plan-Do-Study Act (PDSA) quality improvement process to improve their AFPS quality score.
- The developer also notes that the Bright Futures (American Academy of Pediatrics and the Maternal and Child Health Bureau) guidelines recommended developmental services "include family psychosocial assessment and follow-up, which consists of maternal depression; mental health of parents; smoking, alcohol and drug use; presences of adequate economic, social, and emotional supports; guns; family violence; and other safety issues."
- In the Performance Gap section, the developer reports a HRSA study "study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the Assessment of Family Psychosocial and Well-Being quality measure. Parents were more likely to be asked about one or more psychosocial (family assessment) topics, including mental health and emotional support, at follow-up." The results are included in the testing attachment.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Guidance from the Evidence Algorithm: Patient-reported outcome (Box 1)→ Relationship between PRO and provider action (Box 2)→ Pass

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer reports wide variation across providers who asked parents about all survey items related to family emotional and mental health issues across observed 5 providers. A range of 6.6%-44.3% of parents of young children reported that their child's pediatric clinician discussed psychosocial topics such as parent emotional well-being and partner support in parenting.

Disparities

- The developer states that its analysis of the online PHDS found that variation is observed according to a child's age; race/ethnicity (Hispanic=31.3%, white=23.3%, black=33.0%, Asian=27.3%); level of risk for developmental, behavioral, or social delays across (low/no risk=21.3%, high/moderate risk=29.2%) and respondent education level (did not complete HS=37.1%, HS or higher education=25.5. Non-Hispanic white children are less likely to meet criteria on the Family Assessment measures. Children of lower educated mothers and children at high risk for developmental delay are more likely to have high Family Assessment scores."
- The developer also reports the results of the (Kaiser Permanente Northwest (KPNW) study: "After controlling for other child and family demographic and health factors and provider characteristics, the likelihood (or adjusted odds ratio-AOR) that a child met quality measure criteria differed significantly according to: (1) child's age and (3) child's birth order [Not first born=34.9%, first born=41.7%]."

Question for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

- There is evidence to suggest the patient reported outcome contributes to provider action.
- The questions provided in the attachment do clearly seek parents' report of whether providers raised questions about behavior, development or learning, whether the parents had concerns regarding these things and whether they received information. Responses although influenced by parents' ability to recall are clearly linked to processes of care and high-quality care according to AAP recommendations
- Patient-Centered Outcome measure regarding implementation of national standards for well-child care. Authors presented evidence that practices have used it to drive improvement. Supplements existing measures.
- wide variation across providers.
- Extensive gap analysis and gap/disparities analysis provided. Clear within and between provider variation demonstrated (modelling appropriately adjusted for potential clustering within providers). However, small sample sizes -- not a regional or national sample. However variation unlikely to diminish in larger sample.
- Yes.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability

2a1. Reliability Specifications

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented.

Data source(s): Other – patient/family reported survey

Specifications:

- Level of analysis: Clinician – individual
- Interpretation of score: Better quality = Higher score
- Patient-Reported Outcome-Based Performance Measure (PRO-PM)
- Numerator: *The numerator is the number of parents who had a well child visit within the last 12 months and who were asked about psychosocial well-being.*
- Denominator: *The number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months and all answered questions related to the family psychosocial screening scale (see Attachment A-2, page 18).* [Questions are on pg 36 of Appendix A.]
- There are no exclusions.
- The developer includes a calculation algorithm.
- The measure is not risk adjusted nor risk stratified, but the developer states that it can be stratified by variables such as child demographics characteristics (e.g., the child´s age, race); child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or parent health characteristics, if large enough data sets are available.
- The measure does not use sampling.
- This measure relies on a set of questions within the Promoting Healthy Development Survey (page 36 of the Appendix). This online survey is initiated by the provider who sends it to a parent after a well-child visit. Providers must have a minimum of 10 surveys to generate a report to maintain parent confidentiality.

Questions for the Committee:

- *Are all the data elements (question items) clearly defined? Are all appropriate codes included?*
- *Is the logic or calculation algorithm clear?*
- *Is it likely this measure can be consistently implemented?*

2a2. Reliability Testing Testing attachment

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers.

SUMMARY OF TESTING

Reliability testing level ☐ **Measure score** ☐ **Data element** ☒ **Both**

Reliability testing performed with the data source and level of analysis indicated for this measure ☒ **Yes** ☐ **No**

NQF Note: Both measure score and data element reliability testing are required for PRO-PMs.

Method(s) of reliability testing

- The developer used data from the online Promoting Healthy Development Survey (PHDS), the KPNW study, and a HRSA evaluation study that was testing "three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young children in the context of discussions between pediatric clinicians and parents during well-child visits"; the HRSA study used 5 tools including the PHDS.
- The Cronbach alpha to test internal consistency (data element reliability) was calculated using the Online PHDS and KPNW data. In addition, factor analysis was performed to investigate the dimensionality of the scale.
- To test the PHD survey itself, inter-item correlation was "assessed to insure redundancy of the questions."
- Score level reliability was assessed by inter-unit reliability (IUR) using analysis of variance.
- Intra-class correlation (ICC) was calculated using ANOVA.

Results of reliability testing

- Using the top 5 individual providers with the highest number of surveys (N=77 to 94) from the online PHDS testing, the developer reports the Cronbach's alpha for internal consistency (item-level) range from 0.75-0.95 with the score for all providers at 0.88. The top 5 providers for the KPNW study (N=66 to 80) had Cronbach's alphas ranging from 0.72-0.86, all providers scoring 0.81.
- No data are provided on the inter-item correlation analyses.
- The developer reports that the results for the inter-unit reliability (IUR) testing, at 0.70, which the developer indicates is at on the border of the recommended threshold (0.70) to reliably demonstrate differences between providers. Providers with 20 or more surveys were assessed; no information on the N is provided for these analyses.
    - [*NQF note: IUR measures the proportion of the measure variability that is attributable to the between facility variance. A small IUR (near 0) reveals that most of the variation of the measures between facilities is driven by random noise, indicating the measure would not be a good characterization of the differences among facilities, whereas a large IUR (near 1) indicates that most of the variation between facilities is due to the real difference between facilities. The recommended range is above 0.70.*]
- The ICC was 0.76; according to the literature cited by the developer values above 0.74 are considered excellent.

Questions for the Committee:

- *The developer does not provide information on the size of the sample used for the IUR, except stating that providers with 20 or more survey were assessed. Does the Committee wish to discuss sample size with the developer?*
- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient reliability so that differences in performance can be identified?*

Guidance from the Reliability Algorithm: Precise specifications (Box 1)→ Empirical reliability testing (Box 2) → Score level testing (Box 4) → Appropriate method used (Box 5) → High certainty or confidence that the performance measure scores are reliable (Box 6a)→Moderate

Highest possible rating is HIGH

Note: PRO-PMS *require* element-level testing as well, which was conducted; the highest possible rating for data-element only testing is MODERATE.

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Rationale: The developer reports some results based on top 5 providers (n= 77-94 surveys). The IUR was calculated using providers with 20 or more surveys. From this we have inferred the sample size is likely sufficient for the IUR calculations.

2b. Validity

2b1. Validity: Specifications

2b1. Validity Specifications. This section should determine if the measure specifications are consistent with the evidence.

Specifications consistent with evidence in 1a. ☒ **Yes** ☐ **Somewhat** ☐ **No**

Specification not completely consistent with evidence

Question for the Committee:

- *Are the specifications consistent with the evidence?*

2b2. Validity testing

2b2. Validity Testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.

SUMMARY OF TESTING

Validity testing level ☐ **Measure score** ☐ **Data element testing against a gold standard** ☒ **Both**

Method of validity testing of the measure score:

- ☒ **Face validity**
- ☒ **Empirical validity testing of the measure score**

Validity testing method:

- Factor analysis was conducted to assess the construct validity of the quality measure. A Scree test was used to determine the number of factors to extract. Both oblique and orthogonal rotations were evaluated with promax and varimax methods used, respectively.
- To assess the concurrent validity of the quality measure, hypothesized associations among PHDS items and scales were examined. The developer examined two hypotheses linking the discussion of family psychosocial issues with anticipatory guidance and family centered care.
- Pearson correlation coefficients were calculated across all scale measures to test hypotheses about expected relationships among the PHDS quality measures and to assess the degree to which each of the PHDS quality measures provide unique information. The developer notes that "We expect a moderate or strong correlation between the family psychosocial assessment scale measure (>0.30) and inter-scale correlation coefficients to be less than 0.80."
- For face validity, the PHDS items were tested using focus groups, in-depth cognitive interviews, a literature review, and an advisory board of expert stakeholders.

Validity testing results:

The developer reports the following ranges of results:

- "The lowest average factor loading of 0.49 for the measure is attributable to uniformly low scores with little variation observed across individuals or health plans for the survey items comprising the family assessment quality measure. Factor analysis suggests that the scale items are unidimensional." An appropriate favor loading score is above 0.60.
- The concurrent validity testing results demonstrated improved confidence in protecting children from injury if providers talked about keeping homes and cars safe (odds ratio [OR]: 5.9, 95% confidence interval [CI]: 3.4-

10.2; OR: 8.3, 95% CI: 5-13.8) but also showed that parents report they are rarely asked about other psychosocial issues, including gun safety or how parenting works into their daily activities.

- The developer provides a table of Pearson Correlation Coefficients, which assesses whether the measures are examining different topics. The results suggest, according to the developer, that the measures are not redundant, with an average correlation of 0.34. This measure was most highly associated with the *Assessment of family psychosocial well-being measure* (0.54).

Questions for the Committee:

- *Is the test sample adequate to generalize for widespread implementation?*
- *Do the results demonstrate sufficient validity so that conclusions about quality can be made?*
- *Do you agree that the score from this measure as specified is an indicator of quality?*

2b3-2b7. Threats to Validity

2b3. Exclusions:
N/A

Questions for the Committee:

- *Is the lack of exclusions consistent with the evidence?*

2b4. Risk adjustment: **Risk-adjustment method** ☒ **None** ☐ **Statistical model** ☐ **Stratification**

Conceptual rationale for SDS factors included? ☒ **Yes** ☐ **No**

SDS factors included in risk model? ☐ **Yes** ☒ **No**

Risk adjustment summary

- The developer does not risk adjust the measure because "we do not expect variation in the quality of care provided for children due to risk factors, e.g. children with special health care needs. The provider's performance should be the same regardless of risk factors."
- However, the measure can be stratified by several demographic or health variables as "Identification of variation in quality measures across subgroups of children helps to highlight aspects of care and population of children for which preventive and developmental services may be most need of improvement."
- The developer reports that many studies have shown differences in access to and quality of care, as well as parent satisfaction. The developer states that "One study found: Overall, 94.0% of parents reported 1 or more unmet needs for a number of aspects of care, including assessing family alcohol use, substance abuse and safety. Uninsured children and children aged 18 to 35 months are disproportionately represented among the 15.3% of children whose parents indicated an unmet need this area of care. There are significant variations in performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors."
- Variations were observed by demographic and socioeconomic factors.

Questions for the Committee:

- *Do you agree with the developer's rationale that there is no conceptual basis for adjusting this measure for SDS factors?*

2b5. Meaningful difference *(can statistically significant and clinically/practically meaningful differences in performance measure scores can be identified)*:

- To assess meaningful differences, the developer analyzed the top 5 providers (number of individual surveys completed) in the online PHDS; reported on across 56 providers using KPNW study data; and reported on pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.
  - Online PHDS: For the top 5 individual providers with the highest numbers of surveys (n=77 to 94), a range of 25.3%-54.4% of parents of young children report being asked about at least one of three emotional and mental health issues; for all children the rate was 53.0%.
  - KPNW Study: 38.1% of children had parents reporting that providers asked about at least one emotional and mental health issue. Range across providers in the proportion of children who met quality measure

criteria was substantial and statistically significant, 18.5% to 65.2% (p=0.002). Provider n ranged from 15-153.
    - When the provider was used as the level 2 clustering variable, only 1.1-2.2% of the total variance observed was explained by either measured or unmeasured differences between providers. The developer indicates that this "suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers" and that "the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across the all quality measures" (i.e., providers are inconsistent and going to a different provider may not improve a child's care). However, the HRSA study does demonstrate that providers can improve their performance with an intervention.
- HRSA study: The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the PHDS quality of care measures. Parents were more likely to be asked about one or more psychosocial (family assessment) topics, including mental health and emotional support, at follow-up, across all children's characteristics measured. For example, for children ages 3-9 months, baseline performance was 23.9% and follow up performance was 45.0%.

Question for the Committee:

- *Does this measure identify meaningful differences about quality?*

2b6. Comparability of data sources/methods:

N/A

2b7. Missing Data

The developer reports the following:

- **Online PHDS:** Rate of survey completion was calculated based on survey start and complete dates for each respondent. According to the quality measure scoring protocol, if a parent answered less than half of the items in the AFAUSAS measure, their score is considered to be missing. This does not include items that should have been appropriately skipped. Missing responses are not given a valid score and are not included in the calculation of the quality measure.
- **Online PHDS** data show that 2.6% of parents who started the survey did not complete the survey.
- **KPNW Study:** Of the 5,755 sampled children, 2,173 surveys were returned (37.8%). For these children, the provider the parent identified and the provider to which the child was assigned by the health plan were the same 97.3% of the time. A 95% response rate was obtained for the provider survey.
- The developer notes that responses for the KPNW survey did not differ by gender or insurance type, but did differ by age and by number of previous well visits.
- The developer states that missing data are excluded from the analysis and that surveys in which data are missing for two or more questions on the Assessment for Family Psychosocial Screening scale are excluded from the analysis.
- The developer states information about non-respondents is not available, but "Overall, the quality measure had 2.6% of missing cases, ranging 0-7.8% across the top 5 providers with highest number of surveys. Few overall missing values suggest that the measure level results unlikely to be biased by non-response to the survey questions."

Question for the Committee:

- *This measure has a higher rate of missing data than the other PHDS-derived measures (other measures ranged 0-3.2, 3.3, etc, whereas this one ranged 0-7.8.) Is that a concern?*

Guidance from the Validity Algorithm: Specifications consistent with evidence (Box 1)→ Threats to validity addressed (Box 2) →Empirical validity testing (Box 3)→ Measure score testing (Box 6)→ Appropriate method (Box 7)→ Moderate certainty or confidence that the performance measure scores are a valid indicator of quality (Box 8b)→ Moderate

The highest possible score is MODERATE.

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**
RATIONALE: Missing data are not fully addressed; non-respondent bias not available

Committee pre-evaluation comments

Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2d)

- Appropriate. Questions provided have face validity for assessing the PRO in question. Would request a more direct description of how responses to the 3 questions are used to define the numerator.
- PRO-PM: both data element and score levels. developer reports results based on the top 5 providers. ?moderate reliability.
- Used the inter-unit reliability (IUR) using analysis of variance -- fraction of the variation among provider scores that is due to real differences, rather than due to chance and test re-test with ICC = .76
- The developer does not provide information on the size of the sample used for the IUR, except stating that providers with 20 or more survey were assessed. The developer should be asked to clarify the sample size of the IUR.
- PRO-PM: face validity and empirical validity testing. i do question validity with missing data/only 38% returned surveys.
- "Construct validity was assessed for items used in the PRO. Calculated correlation among different items on the survey with usually fair correlation"
- No risk adjustment -- agree this is appropriate. Also agree that stratification by both risk of developmental & behavioral problems as well as by demographic factors can illuminate disparities
- Meaningful differences shown although the sample for assessment is modest
- Impact of missingness not robustly assessed due to lack of data on non-responders"
- Missing data rate is is 7.8%, which may limit ability to identify differences between sites due to noise, but does not invalidate measure.

Criterion 3. Feasibility

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- This is a patient reported outcome. Data are generated by parents completing the CAHMI-developed Promoting Healthy Development Survey (PHDS), which is sent to them by their provider following a well child visit.
- Although the survey has been in use since 2001, there is not currently an automated reporting system for providers. The developer has been working on a new website for the survey that will automatically report data, and expects it to launch in February 2017.
- The developer reports that the provider registration takes about 10 minutes and the parent survey takes about 15-20 minutes. There are no fees, licensing requirements, etc., to use the measure.

Questions for the Committee:

- *Is the data collection strategy ready to be put into operational use?*
- *Does the developer have a status update on the new website?*

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments

Criteria 3: Feasibility

- This measure seems burdensome; parent survey takes 20 minutes; and i'm not convinced of this being an optimal method of improving patient care, experience, or outcome.
- Overall response rate appeared to be about 40%. Those that start the survey tend to finish it. This is a consistent limitation of survey data that is collected outside of the provision of care.
- Since this data is collect via survey, it requires work by the provider to be implemented. Other patient experience surveys are already in place due to JCAHO and other regulatory bodies- this "niche" survey may be challenging to implement in that environment, if we are to avoid survey-fatigue.

Criterion 4: Usability and Use

4. Usability and Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

Current uses of the measure
Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No** ☐ **UNCLEAR**
OR
Planned use in an accountability program? ☐ **Yes** ☒ **No**

Accountability program details
No confirmed use for an accountability program, but CAHMI has been in discussion with a number of organizations that are interested in using the measure, including CMS/Medicaid, Title V, and Head Start.

Improvement results
The developer provided the following response: "In a 2010-2012 study of a large pediatric practice in Oregon (n=551 providers), family assessment increased from 21.5% (n=116) at baseline (2010) to 43.7% (n=111, 2011-12, AOR=3.32, 2.24-4.91) post implementation of the CAHMI Well Visit Planner - a family engagement tool to assist parents in planning for their well child visit. This represents a 103.3% increase and was statistically significant at the 95% confidence level. The PHDS, which contains Assessment of Family Psychosocial Screening measure, was used as the evaluation tool."

Unexpected findings (positive or negative) during implementation
The developer was not aware of any unintended consequences.

Potential harms
The developer was unaware of any potential harms.

Vetting of the measure
The developer conducted key informant interviews and focus groups with patients and providers during testing. The developer reports that "The feedback was helpful for future implementation efforts of CAHMI's family engagement tools. The feedback, however, did not result in any changes to the measure itself." The developer also reports positive feedback on the measure from both patients and providers.

Feedback:

N/A

***Questions for the Committee**:*
- o *Can the performance results be used to further the goal of high-quality, efficient healthcare?*
- o *Do the benefits of the measure outweigh any potential unintended consequences?*

Preliminary rating for usability and use: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee pre-evaluation comments
Criteria 4: Usability and Use

- Assessment of psychosocial wellbeing is an important component of quality patient care, but i don't think there is enough strong information to say asking families for feedback on if they were asked about it equates to improving patient care in a meaningful way.
- Can developers consider sharing aggregate results with patients who participate so that they might see providers' performance on the measure across all patients?

- Used primarily in improvement work at present. No public reporting done.

Criterion 5: Related and Competing Measures

Related or competing measures

This measure is part of a set of five based on the PHD survey:

- 3219: Anticipatory Guidance and Parental Education
- 3220: Ask About Parental Concerns
- 3221: Family Centered Care
- 3222: Assessment of Family Alcohol Use, Substance Abuse and Safety
- 3223: Assessment of Family Psychosocial Screening

Harmonization

N/A

Endorsement + Designation

The "Endorsement +" designation identifies measures that exceed NQF's endorsement criteria in several key areas. After a Committee recommends a measure for endorsement, it will then consider whether the measure also meets the "Endorsement +" criteria.

This measure is a candidate for the "Endorsement +" designation IF the Committee determines that it: meets evidence for measure focus without an exception; is reliable, as demonstrated by score-level testing; is valid, as demonstrated by score-level testing (not via face validity only); and has been vetted by those being measured or other users.

Eligible for Endorsement + designation: ☐ **Yes** ☒ **No**

RATIONALE IF NOT ELIGIBLE: The measure has not been vetted by those being measured or other users.

Pre-meeting public and member comments

- None

NATIONAL QUALITY FORUM—Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2970
Measure Title: Assessment of Family Psychosocial Screening
IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title
Date of Submission: 1/13/2017

Instructions

- *Complete 1a.1 and 1a.12 for all measures.*
- *Complete **EITHER 1a.2, 1a.3 or 1a.4** as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus
The measure focus is evidence-based, demonstrated as follows:

- Health outcome: [3] a rationale supports the relationship of the health outcome to processes or structures of care. Applies to patient-reported outcomes (PRO), including health-related quality of life/functional status, symptom/symptom burden, experience with care, health-related behavior.
- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: [5] a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence [4] that the measured structure leads to a desired health outcome.
- Efficiency: [6] evidence not required for the resource use component.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.
4. The preferred systems for grading the evidence are the U.S. Preventive Services Task Force (USPSTF) grading definitions and methods, or Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines.
5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.
6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)
Outcome
☒ Health outcome: Click here to name the health outcome
☒Patient-reported outcome (PRO): Assessment of Family Psychosocial Screening (AFPS)

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Figure 1 (attached) shows the logic model by which the Assessment of family Psychosocial Screening (AFPS) measure is obtained and improved. Simply said: (1) the parent and child attend a well child visit with their provider; (2) the provider subsequently sends a survey -- the Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org), which includes one question (3 items, see Attachment A-2, page18) for the parent to complete; (3) when at least ten surveys have been completed, the provider receives a feedback report on parents' experiences of the visit and the extent to which they received an assessment of family psychosocial screening via the CAHMI PHDS Toolkit website (www.phdstoolkit.org); (4) the provider reviews the report and then can engage in a *Plan-Do-Study Act* (PDSA) quality improvement process to improve their AFPS quality score. THE PDSA cycle involves reviewing the baseline data; developing and implementing a plan of action to improve the score; obtaining further data from the parent; and comparing the first set of results with the second. The full process is repeated until providers are satisfied with their improved scores. We are currently applying for this process to be approved by the American Board of Pediatrics (ABP) for maintenance of certification (MOC, Part 4) credit. The provider must complete three PDSA cycles. Each time point must have at least 25 completed surveys and there must be at least 8 weeks between time periods.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

Recommended developmental services, as set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau, include family psychosocial assessment and follow-up, which consists of maternal depression; mental health of parents; smoking, alcohol and drug use; presences of adequate economic, social, and emotional supports; guns; family violence; and other safety issues.[1] In order to gauge the quality of recommended care provided, this type of information must be collected from the parent in order to identify the level at which providers discuss these issues with parents. Previous studies have shown that parents are willing to discuss such sensitive topics with providers.

Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child's health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews which are not accurate for the specific level of information obtained in the Promoting Healthy Development Survey (PHDS, see Attachment A-2, page 18). The process outlined in the logic model (1a.12) allows health care providers to better understand the extent to which their patients experience "quality care" – in this case, the extent to which parents received assessment of family psychosocial screening. It also allows providers to engage in quality improvement activities to improve their parent-reported Family Centered Care quality score by using several Plan-Do-Study Act (PDSA) cycles, as described above.

[1] *Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents*, 3rd Edition

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

.

1. Evidence, Performance Gap, Priority – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

Figure_1_Family_Psychosocial_Screening_Logic_Model.docx,CAHMI_Psychosocial_Screening_evidence_attachment_revised_02_02_17.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Please update any changes in the evidence attachment in red. Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. If there is no new evidence, no updating of the evidence information is needed.

No

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. **Briefly explain the rationale for this measure** *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*
IF a PRO-PM (e.g. HRQoL/functional status, symptom/burden, experience with care, health-related behaviors), provide evidence that the target population values the measured PRO and finds it meaningful. (Describe how and from whom their input was obtained.)
IF a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and provide rationale for composite in question 1c.3 on the composite tab.

Recommended developmental services, as set forth by the American Academy of Pediatrics (AAP) and the Maternal and Child Health Bureau, include family psychosocial assessment and follow-up, which consists of maternal depression; mental health of parents; smoking, alcohol and drug use; presences of adequate economic, social, and emotional supports; guns; family violence; and other safety issues. In order to gauge the quality of recommended care provided, this type of information must be collected from the parent in order to identify the level at which providers discuss these issues with parents. Previous studies have shown that parents are willing to discuss such sensitive topics with providers.

Few standardized quality measures are available that provide specific information about preventive health care for young children, especially on aspects of care for which parents and families are a reliable source of information about the quality of their child´s health care. A majority of the measures currently used provide information about whether children come in for well-child visits (access to care measures) or are based on medical chart reviews which are not accurate for the specific level of information obtained in the Promoting Healthy Development Survey (PHDS).

1b.2. **Provide performance scores on the measure as specified (current and over time) at the specified level of analysis**. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*

see also Testing Form-Family Psychosocial Assessment

DATA SOURCES

Differences in the quality measure scores across providers is demonstrated for (1) 5 top individual providers with the highest number of surveys using Online PHDS data; (2) across 56 providers using Kaiser Permanente NW study data; and (3) pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.

Online PHDS: The performance scale for the quality measure was calculated using the scoring methods described in the Attachment . Individual provider level differences in performance were illustrated by the proportion of children meeting the quality of care criteria across 5 top providers with the highest number of completed surveys after their well-child visit.

KPNW Study: The significance of differences observed in the proportion of children meeting criteria for the quality measure across pediatric providers (n=56) was evaluated using t-tests. The relative spread in the quality measure score across providers was assessed using the coefficient of variation statistics (standard deviation across providers multiplied by 100%). Multi-level regression models were conducted using the pediatric provider as the level 2 clustering variable, in order to assess the degree to which the probability that a child meets criteria on each quality measure is explained by differences between providers (called the "clustering effect"). In implementing this multi-level regression method (Empty Model), the presence of a significant clustering effect by pediatric providers was estimated prior to accounting for the child and family characteristics associated with each provider. Second, variables related to the child and family characteristics (child's age, gender, race/ethnicity, birth order, developmental and behavioral delay risk status; parent education and risk for depression) were added to the Empty Model to assess how much of the provider clustering effect observed remains after accounting for these characteristics (called the "Patient Model").

HRSA study: Quantitative data results for the baseline (2010) and follow-up (2011-12) study of the intervention sites using the HRSA Evaluation Study data were conducted using basic descriptive statistics to describe each sample and applying chi-square test of statistical significance to assess differences in the quality measure for the baseline and follow-up samples.

PERFORMANCE RESULTS

Online PHDS:

Table 1b.2a present the proportion of children whose care met for the quality measure across 5 providers. Variation across providers who asked parents about all survey items related to family emotional and mental health issues is substantially wide across observed 5 providers. Only 6.6%-44.3% of parents of young children reported that their child's pediatric clinician discussed psychosocial topics such as parent emotional well-being and partner support in parenting.

Table 1b.2a: Proportion of Children Meeting Measure Criteria, Top 5 individual providers with highest number of surveys

Characteristics (n=5355)	All Children				
	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
	1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
Asking parents about at least one of three emotional and mental health issues (at least one of these topics)	53.0%	39.1%	25.3%	51.8%	54.4% 47.3%
Asking parents about at least one of three emotional and mental health issues (asked all items)	27.0%	28.3%	6.6%	37.3%	44.3% 29.7%

KPNW Study: 38.1% reported that providers discussed at least one emotional and mental health topics. Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant (p=0.002)

Table 1b.2b: Proportion of all children in the study who met criteria for receiving quality services and ranges in proportion across providers. (SD=Standard Deviation)

Developmental Services Quality Measures (n = 2173)	Proportion of All Children Meeting Measure Criteria	Range in the Proportion of Children Meeting Measure Criteria Across 51 Pediatric Providers	Relative Variation (COV) in Measure Scores Across Pediatric Providers
Asking parents about at least one of three emotional and mental health issues (at least one of these topics)	38.1%	18.5% to 65.2%	SD: 10%; (p = 0.002) 26.1%

Only providers with n=15 or more PHDS responses are included in the provider level analysis. Provider level n ranges from 15 to 153.

Multi-level analysis: For the Empty Model that used the provider as the level 2 clustering variable, only 1.1% to 2.2% of the total variance observed in whether children met criteria for each quality measure was explained by either measured or unmeasured differences between the providers that they see. This suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers. These findings translate into a 1.19 to 1.29 median odds ratio across the six quality measures, indicating that the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across the all quality measures. When child/family level

characteristics are added to the model (Patient Model), the total variance explained by differences between providers does not change significantly.

HRSA study
The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the Assessment of Family Psychosocial and Well-Being quality measure. Parents were more likely to be asked about one or more psychosocial (family assessment) topics, including mental health and emotional support, at follow-up. The tables below present comparison of percent of children who received care met the quality care criteria between baseline and follow-up survey data.

Table 1b.2c. Family Assessment*, by Children's Characteristics: ---
Parent was asked about one or more family assessment topics

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test p value
Age			
3-9 months	23.9% (132)	45.0% (145)	<0.0001
10-18 months	21.5% (87)	34.4% (111)	<0.0001
19-48 months	29.3% (103)	50.5% (112)	<0.0001
Race			
Hispanic	26.0% (26)	47.8% (22)	0.01
White	24.5% (252)	42.7% (305)	<0.0001
Asian	28.6% (8)	(4)	-
Multiple or other	50.0% (12)	50.0% (9)	1.00
Insurance type			
Private or private and public	22.9% (248)	37.9% (257)	<0.0001
Public only (includes Medicaid, Medicare, CHIP, and Military)	33.5% (64)	39.1% (63)	<0.0001
Other insurance type	(2)	(1)	N/A
Uninsured	(4)	(3)	N/A
At risk of developmental delay			
Low/no risk	24.3% (248)	41.6% (227)	<0.0001
High/moderate risk	26.0% (73)	44.7% (76)	<0.0001

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.
NA

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.*
See also Testing Form-Family Psychosocial Assessment
DATA SOURCES
We used the following data sources for testing of the quality measure:
(1) Online Promoting Healthy Development Survey (PHDS) – data collected through an online, publicly available tool (Promoting Healthy Development Survey-PHDS). Parents who had a well-child care visit in the last 12 months can complete the PHDS. Providers initiate the survey. (See Evidence Form Figure 1 for visual model of the Online PHDS.)
2) Kaiser Permanente Northwest (KPNW) Study – CAHMI partnered with Kaiser Permanente Northwest in Portland, Oregon. The study aimed to evaluate the level and variations in the quality of preventive and developmental services for young children and assess the contribution of key system, provider and patient factors.

STUDY POPULATION

Online PHDS: Children age 3-48 months of age whose parents completed the online publicly available PHDS were included in the testing. Between 2008-2016, we received 5,670 completed surveys. Of those surveys, 5,355 surveys with provider IDs were used for analyses. Children's socio-demographic and health characteristics varied across the individual providers included in the analysis.

Table 1b.4a: Characteristics of children for whose visited provider ID is available

Characteristics (n=5355)	All Children	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)					
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)	
Age of child							
Under 10 months of age	38.3%	19.1%	49.5%	33.3%	54.4%	24.7%	
10 to 18 months of age	34.7%	39.4%	38.5%	38.9%	29.1%	57.1%	
19-47 months of age	27.0%	41.5%	12.1%	27.8%	16.5%	18.2%	
Race/ethnicity of child							
White, non-Hispanic	53.8%	13.3%	81.0%	20.3%	50.7%	17.3%	
Hispanic	40.8%	81.1%	14.3%	74.7%	40.8%	78.7%	
Other race/ethnicity	5.3%	5.5%	4.7%	7.0%	8.4%	4.0%	
Respondent education level							
Did not complete high school	12.1%	23.6%	0	34.1%	6.4%	15.8%	
Completed high school	88.9%	76.4%	100%	65.9%	93.6%	84.2%	
Children who qualify for Children with Special Health Care Needs (CSHCN) Screener criteria							
CSHCN	10.1%	7.4%	8.8%	10.0%	11.4%	5.2%	
Non-CSHCN	89.9%	92.6%	91.2%	90.0%	88.6%	94.8%	
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	22.7%	-	24.4%	-	28.9%	0%	

-Data is not available due to small sample size

KPNW Study: The population studied was children 3 to 48 months old who live in a metropolitan area in the Pacific Northwest. One randomly selected child per household whose age would be no younger than 3 months of age and no older than 48 months of age at the time that their parents received the survey and had one or more well-child visits were eligible to be sampled. A random sample of 5,755 children were identified. Of the 5,755 sampled children, 2,173 surveys were returned (37.8%).

Table 1b.4b: Characteristics of children for whom survey responses were received, KPNW study, Top 5 individual providers with highest number of surveys

Characteristics (n=2173)	All Children	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)					
		7 (n=80)	53 (n=77)	4 (n=74)	1 (n=67)	43 (n=66)	
Age of child							
Under 10 months of age	22.0%	20.0%	19.5%	20.3%	22.4%	21.2%	
10 to 18 months of age	26.6%	25.0%	29.9%	35.1%	22.4%	15.2%	
19-47 months of age	51.4%	55.0%	50.6%	44.6%	55.2%	63.6%	
Gender of child							
Female child	46.2%	48.8%	49.4%	47.3%	41.8%	45.5%	
Male child	53.8%	51.3%	50.6%	52.7%	58.2%	54.5%	
Race/ethnicity of child							
White, non-Hispanic	72.9%	84.8%	77.0%	93.2%	76.9%	62.5%	
Asian, non-Hispanic	7.8%	2.5%	6.8%	1.4%	3.1%	20.3%	
Hispanic	8.9%	6.3%	12.2%	2.7%	10.8%	10.9%	
Other race/ethnicity	10.4%	6.3%	4.1%	2.7%	9.2%	6.3%	
Child is the first born in the family	52.1%	52.5%	40.8%	35.1%	54.5%	52.3%	
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	31.3%	21.5%	24.7%	27.0%	29.7%	26.2%	
Education level of mother							
High school or less	12.7%	20.3%	3.9%	14.9%	16.7%	6.2%	

More than high school	87.3%	79.7%	96.1%	85.1%	83.3%	93.8%

DISPARITIES

Online PHDS: Variation is observed according to a child's age; race/ethnicity; level of risk for developmental, behavioral, or social delays across and respondent education level. Non-Hispanic white children are less likely to meet criteria on the Family Assessment measures. Children of lower educated mothers and children at high risk for developmental delay are more likely to have high Family Assessment scores.

Table 1b.4d. Assessment of family psychosocial well-being (asked about all items) by child demographic and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	779	34.7%
9-18 months	432	22.9%
19-48 months	279	20.1%
p values (Pearson chi-square)	-	<0.0001
Gender		
Male	188	30.3%
Female	191	30.5%
p values (Pearson chi-square)	-	0.96
Race/ethnicity		
Hispanic	610	31.3%
White non-Hispanic	664	23.3%
Black non-Hispanic	33	33.0%
Asian non-Hispanic	30	27.3%
Other/Multi race, non-Hispanic	24	28.2%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	227	37.1%
Completed high school or higher education	1213	25.5%
p values (Pearson chi-square)	-	<0.0001
CSHCN status		
Non-CSHCN	1333	26.9%
CSHCN	157	27.6%
p values (Pearson chi-square)	-	0.71
At risk for developmental delay (online only)		
Low/No risk	476	21.3%
High/Moderate risk	216	29.2%
p values (Pearson chi-square)	-	<0.0001

KPNW study: After controlling for other child and family demographic and health factors and provider characteristics, the likelihood (or adjusted odds ratio-AOR) that a child met quality measure criteria differed significantly according to: (1) child's age and (3) child's birth

Table 1b.4d: Mean number of developmental services care components for which quality care was received and the proportion of children meeting criteria for receiving quality developmental services by characteristics of children and families.

Characteristic of Child or Child's Family	% Meeting PSYCH Criteria
Child's Age	
Less than 9 mos.	56.4% S
10 to 18 mos.	37.1%
AOR: .47	
19 to 49 mos.	30.8%
AOR: .35	
Child's Gender	
Male Child	37.8% NS

Female Child 38.5%
Child's Race
White, Non-Hispanic 38.5% NS
Asian, Non-Hispanic 42.4%
Hispanic 37.8%
Other Race,
Multiple Race 36.9%
Birth Order
Not First Born 34.9% S
First Born 41.7%
AOR: 1.29
Child's Risk for Developmental, Behavioral or Social Delays (Using Parent's Evaluation of Developmental Status)
Low/No Risk 37.1% NS
At Risk 39.7%
Respondent Education
More than High School 38.0% NS
High School or Less 40.9%
Respondent's Risk for Depression (Using the Kemper Screener)
No Symptoms of Depression 37.8% NS
Symptoms of Depression 41.3%
NOTE: Adjusted odds ratios (AOR) derived from regression analyses listed in the table are shown only if they are statistically significant. AOR uses the first subgroup of each characteristic as a reference. S differences significant at the p < .05 level of significance; NS differences not significant.

1b.5. **If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4**
NA

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. **Subject/Topic Area** *(check all the areas that apply):*

De.6. **Cross Cutting Areas** *(check all the areas that apply):*
«crosscutting_area»

De.7. **Target Population Category** *(Check all the populations for which the measure is specified and tested if any):*

S.1. **Measure-specific Web Page** *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*
NA

S.2a. **If this is an eMeasure**, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. **Data Dictionary, Code Table, or Value Sets** *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*
Attachment **Attachment:** CAHMI_Data_Dictionary_Assessmen_Family_Psychosocial_Screening-636203523677481095.pdf

S.3.1. **For maintenance of endorsement:** Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.
No

S.3.2. **For maintenance of endorsement,** please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.
NA

S.4. **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*
IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator is the number of parents who had a well child visit within the last 12 months and who were asked about psychosocial well-being.

S.5. **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*
IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The numerator is the number of parents who had a well child visit within the last 12 months and who were asked about psychosocial well-being.

S.6. **Denominator Statement** *(Brief, narrative description of the target population being measured)*
The number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months and all answered questions related to the family psychosocial screening scale (see Attachment A-2, page 18).

S.7. **Denominator Details** *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).
The number of parents with children ages 0-48 months who have completed a well child visit within the last 12 months and all answered questions related to the family psychosocial screening scale (see Attachment A-2, page 18).

S.8. **Denominator Exclusions** *(Brief narrative description of exclusions from the target population)*
Missing data are excluded from the analysis.

S.9. **Denominator Exclusion Details** *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*
Surveys in which data are missing for two or more questions on the Assessment for Family Psychosocial Screening scale are excluded from the analysis. Approximately 2.6% of parents who started the Online PHDS did not complete the survey (range 0.0-7.8% for top 5 providers with highest number of surveys; see Testing form, pages 21-23 for more detailed information on missing data).

S.10. **Stratification Information** *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and*

coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)
Although no stratification is required, the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the findings by possible vulnerability, should any individual provide have sufficient data (parent responses) to do so. Potential variables for stratification include:
(1) Child demographic characteristics (e.g., the child´s age, race);
(2) Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs); and/or
(3) Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

S.11. **Risk Adjustment Type** (Select type. Provide specifications for risk stratification in measure testing attachment)
No risk adjustment or risk stratification
If other:

S.12. **Type of score:**
Rate/proportion
If other:

S.13. **Interpretation of Score** *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*
Better quality = Higher score

S.14. **Calculation Algorithm/Measure Logic** (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*
The numerator is the sum of survey respondents (parents) answering "Yes" to two or more of the Assessment of Family Pscyhosocial Screening questions. The denominator is the sum of all respondents answering two or more of the assessment questions. Surveys missing two or more responses to this set of questions are excluded from analysis. An aggregated score of 100% is required for achieving quality for this measure.

S.15. **Sampling** *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*
IF a PRO-PM, identify whether (and how) proxy responses are allowed.
NA

S.16. **Survey/Patient-reported data** *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*
IF a PRO-PM, specify calculation of response rates to be reported with performance measure results.
Data are collected using the parent-reported "Promoting Healthy Development Survey" (PHDS) developed by the CAHMI (www.wellvisitsurvey.org). Instructions for survey completion are included with the survey. Family Centered Care questions are multiple choice (Yes Definitely, Yes, Somewhat, and No). The PHDS is survey is initiated by the provider who can send it to all parents who have received a well child visit. CAHMI has a website (www.phdstoolkit.org) where providers can register to use for the PHDS. This site assigns each provider a unique URL, which allows for provider identification by CAHMI as well as light branding with the provider´s logo so that it is identifiable by the parent. The PHDS Toolkit website sends an email to the provider with the unique URL link to the survey. The provider then sends the link to the parents asking them with instructions to fill out the survey and provide feedback about the visit. The parent fills out the survey and receives a customized feedback report. The survey data are captured on a secure HIPAA compliant CAHMI server. Through the PHDS Toolkit website, providers can generate a report that aggregate parent data information from the survey. Providers must have a minimum of 10 surveys to generate a report to maintain parent confidentiality. See Evidence Form, Figure 1 for a visual model this process.

S.17. **Data Source** *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*
If other, please describe in S.18.
Other

S.18. **Data Source or Collection Instrument** *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data is collected.)*
IF a PRO-PM, identify the specific PROM(s); and standard methods, modes, and languages of administration.

The Assessment of Family Psychosocial Screening quality measure is included as part the CAHMI Promoting Healthy Development Survey (PHDS, www.wellvisitsurvey.org). The data are generated by parents filling out the PHDS. The PHDS is based in English. See Evidence Form, Figure 1 for a description visual model of the data collection process.

S.19. **Data Source or Collection Instrument** *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*
Available in attached appendix at A.1

S.20. **Level of Analysis** *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*
Clinician : Individual

S.21. **Care Setting** *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*
Clinician Office/Clinic
If other:

S.22. **COMPOSITE Performance Measure** - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*
NA

2. Validity – See attached Measure Testing Submission Form
CAHMI_NQF_testing_attachment_Family_Psychosoc_Assessment_020217.docx

2.1 For maintenance of endorsement
Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)
Yes

2.2 For maintenance of endorsement
Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. (Do not remove prior testing information – include date of new information in red.)
No

2.3 For maintenance of endorsement
Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes SDS factors is no longer prohibited during the SDS Trial Period (2015-2016). Please update sections 1.8, 2a2, 2b2, 2b4, and 2b6 in the Testing attachment and S.14 and S.15 in the online submission form in accordance with the requirements for the SDS Trial Period. NOTE: These sections must be updated even if SDS factors are not included in the risk-adjustment strategy. If yes, and your testing attachment does not have the additional questions for the SDS Trial please add these questions to your testing attachment:

What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care)

What were the statistical results of the analyses used to select risk factors?

Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)
No - This measure is not risk-adjusted

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (*if previously endorsed*)**:** 2970
Measure Title: Assessment of Family Psychosocial Well-Being
Date of Submission: 2/2/2017
Type of Measure:

☒ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process	☐ Efficiency
☐ Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- **For outcome and resource use measures**, section **2b4** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; [12]
AND
If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful [16] differences in performance**;
OR
there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For **eMeasures, composites, and PRO-PMs** (or other measures susceptible to missing data), analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.23*)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ administrative claims	☐ administrative claims
☐ clinical database/registry	☐ clinical database/registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record

☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: Patient reported data	☒ other: Patient reported data

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).
We used the following data sources for testing of the quality measure:

1) Online Promoting Healthy Development Survey (PHDS) – data collected through an online, publicly available tool (Promoting Healthy Development Survey-PHDS). Parents who had a well-child care visit in the last 12 months can complete the PHDS. Providers initiate the survey. (See Evidence Form Figure 1 for the Online PHDS logic model.)

2) Kaiser Permanente Northwest (KPNW) Study – CAHMI partnered with Kaiser Permanente Northwest in Portland, Oregon. The study aimed to evaluate the level and variations in the quality of preventive and developmental services for young children and assess the contribution of key system, provider and patient factors.

3) HRSA Evaluation Study - The specific goal of this study was to evaluate the feasibility, acceptability and impact of three different patient-centered strategies for improving the quality and equity of preventive and developmental services provided to young children in the context of discussions between pediatric clinicians and parents during well-child visits. The evaluation measures used data from 5 different tools/surveys including PHDS. The parent-completed PHDS was administered before and after the intervention to assess changes in the quality of well-child care. The study funded by Health Resources and Services and Administration's (HRSA) Maternal and Child Health Bureau. (Patient Centered Quality Improvement of Well-Child Care, Final Report, Supported by a grant from the Maternal and Child Health Bureau Research Grants Program, Health Resources and Services Administration, R40 MC08959 03-00.)

1.3. What are the dates of the data used in testing? 2004-2016
Online PHDS: 2008-2016
KPNW Study: 2004-2005
HRSA Evaluation Study: 2010-2012

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.26*)	**Measure Tested at Level of:**
☒ individual clinician	☒ individual clinician
☐ group/practice	☐ group/practice
☐ hospital/facility/agency	☐ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Online PHDS: n=5,670 surveys reporting on quality of care provided by 299 individual pediatricians and primary care providers from 88 clinics in 36 states. Participation is a voluntary self-selection process based on knowledge and interest in quality improvement in their practice.

KPNW Study: Provider-level surveys and quality of care assessment were focused on the care provided by 56 individual providers (44 pediatricians, 9 nurse practitioners, 3 physician assistants) in the pediatrics department who were organized into ten geographically distinct offices.

HRSA Evaluation Study: Three pediatric offices in Oregon: 1) a rural site, (4 pediatricians), 2) an urban site (8 pediatricians), and 3) an urban site, (12 pediatricians). All pediatricians in selected clinic and office staff participated in relevant baseline and follow up data collection.

1.6. How many and which patients were included in the testing and analysis (by level of analysis and data source)?
(*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Online PHDS: Children age 3-48 months of age whose parents completed the online publicly available PHDS were included in the testing. During 2008-2016, we received 5,670 completed surveys. Of those surveys, 5,355 surveys with provider IDs were used for analyses. Children's socio-demographic and health characteristics varied across the individual providers included in the analysis.

Table 1.6a: Characteristics of children for whose visited provider ID is available

Characteristics	All Children (n=5355)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
Age of child						
Under 10 months of age	38.3%	19.1%	49.5%	33.3%	54.4%	24.7%
10 to 18 months of age	34.7%	39.4%	38.5%	38.9%	29.1%	57.1%
19-47 months of age	27.0%	41.5%	12.1%	27.8%	16.5%	18.2%
Race/ethnicity of child						
White, non-Hispanic	53.8%	13.3%	81.0%	20.3%	50.7%	17.3%
Hispanic	40.8%	81.1%	14.3%	74.7%	40.8%	78.7%
Other race/ethnicity	5.3%	5.5%	4.7%	7.0%	8.4%	4.0%
Respondent education level						
Did not complete high school	12.1%	23.6%	0	34.1%	6.4%	15.8%
Completed high school	88.9%	76.4%	100%	65.9%	93.6%	84.2%
Children who qualify for Children with Special Health Care Needs (CSHCN) Screener criteria						
CSHCN	10.1%	7.4%	8.8%	10.0%	11.4%	5.2%
Non-CSHCN	89.9%	92.6%	91.2%	90.0%	88.6%	94.8%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	22.7%	-	24.4%	-	28.9%	0%

-Data is not available due to small sample size

KPNW Study: The population studied was children 3 to 48 months old who live in a metropolitan area in the Pacific Northwest. One randomly selected child per household whose age would be no younger than 3 months of age and no older than 48 months of age at the time that their parents received the survey and had one or more well-child visits were eligible to be sampled. A random sample of 5,755 children were identified. Of the 5,755 sampled children, 2,173 surveys were returned (37.8%).

Table 1.6b: Characteristics of children for whom survey responses were received, KPNW study, Top 5 individual providers with highest number of surveys

Characteristics	All Children (n=2173)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		7 (n=80)	53 (n=77)	4 (n=74)	1 (n=67)	43 (n=66)
Age of child						
Under 10 months of age	22.0%	20.0%	19.5%	20.3%	22.4%	21.2%
10 to 18 months of age	26.6%	25.0%	29.9%	35.1%	22.4%	15.2%
19-47 months of age	51.4%	55.0%	50.6%	44.6%	55.2%	63.6%
Gender of child						
Female child	46.2%	48.8%	49.4%	47.3%	41.8%	45.5%
Male child	53.8%	51.3%	50.6%	52.7%	58.2%	54.5%
Race/ethnicity of child						
White, non-Hispanic	72.9%	84.8%	77.0%	93.2%	76.9%	62.5%
Asian, non-Hispanic	7.8%	2.5%	6.8%	1.4%	3.1%	20.3%
Hispanic	8.9%	6.3%	12.2%	2.7%	10.8%	10.9%
Other race/ethnicity	10.4%	6.3%	4.1%	2.7%	9.2%	6.3%
Child is the first born in the family	52.1%	52.5%	40.8%	35.1%	54.5%	52.3%
Child has moderate or high risk for developmental, behavioral or social delays (PEDS)	31.3%	21.5%	24.7%	27.0%	29.7%	26.2%
Education level of mother						
High school or less	12.7%	20.3%	3.9%	14.9%	16.7%	6.2%
More than high school	87.3%	79.7%	96.1%	85.1%	83.3%	93.8%

HRSA Evaluation Study: The study inclusion criteria were used to determine which parents/guardians of children were invited to participate in the interventions and/or evaluation from each participating study site:

- Parent has a well-child visit scheduled at this intervention site for one or more of their children.
- The child is scheduled for their 4-month to 3-year-old well-child visit and, therefore, is between the ages of 4 and 40 months (e.g. 40 month old children could be there for their 3 year well-child visit)
- The parent can read and understand English and is able to complete the intervention and evaluation tools.
- For intervention, the parent was able to access the online version of the Plan My Child's Well-Visit tool and the online evaluation survey.

The analysis includes 551 completed surveys at baseline (2010) and 275 completed surveys at follow-up (2011-12)

Table 1.6c. Sample description for baseline and follow-up PHDS respondents

	Baseline (n=551)	Follow-up (n=275)
Visit type of child for whom survey was completed		
4, 6 or 9-month	38.9%	36.2%
12, 15 or 18-month	33.7%	41.3%
24 or 36-month	27.4%	22.4%
Birth order of child for whom survey was completed		
First child	42.2%	56.6%
Not first child	57.8%	43.4%
Race/ethnicity		
White, non-Hispanic	80.3%	83.5%
Hispanic	8.4%	6.6%
Other/multiple, non-Hispanic	8.6%	6.6%
Asian, non-Hispanic	2.7%	3.3%
Insurance type		
Private or private and public	90.7%	86.7%
Public only (includes Medicaid, Medicare, CHIP and Military)	7.6%	12.1%
Other	0.7%	0.4%
None	0.9%	0.8%

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

Online PHDS and KPNW study data were used for reliability testing and stratification analysis. Validity findings are presented from a peer-reviewed publications and online PHDS and KPNW study data. Performance analysis was conducted using the online PHDS, KPNW study and HRSA Evaluation Study data.

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Online PHDS: Child's age, sex, race/ethnicity, and respondent (parent) age, race/ethnicity, and education level. The survey does not have a question asks about family income due to complexity of collecting income data by self-reported survey. However, the online PHDS has items assessing the family's economic situation: How much trouble does the family have paying for a) child's health and medical expenses; b) supplies like formula, food, diapers, clothes and shoes; and c) health care for the parent.
KPNW Study: Child's age, sex, race/ethnicity, and education level of mother
HRSA Study: Child's age, race-ethnicity, and insurance type

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Internal consistency: Cronbach's alpha was calculated using the Online PHDS and KPNW data. Cronbach's alpha is the most widely used in health care research when multiple-item measures of a concept or construct are employed. The acceptable values of alpha ranges from 0.70 to 0.95. In addition, factor analysis was performed to investigate the dimensionality of the scale.
Inter-item correlation was assessed to insure redundancy of the questions.

The primary aim of the Assessment of Family Psychosocial Screening (AFPS) quality measure is to detect difference between providers on whether providers assess the family for psychosocial issues, such as depression and having emotional support. Provider level reliability was assessed by inter-unit reliability (IUR) using analysis of variance. IUR can be interpreted as the fraction of the variation among provider scores that is due to real differences, rather than due to chance. If the IUR is higher, the ability of the item or scale measure to discriminate across programs is greater. Scales with reliability coefficients above 0.70 provide adequate precision for use in statistical analysis of unit-level comparisons.[1] As the IUR gets smaller, a larger sample is needed in order to reliably discriminate across programs. In the analysis we included providers with 20 or more completed surveys.

Intra-class correlation (ICC) was calculated using ANOVA, as a ratio of the variance between groups over the total variance. The interpretation of the ICC is as the proportion of relevant variance that is associated with differences among measured objects.[2] Fleiss (1981) and Cicchetti and Sparrow (1981) from the medical group state that ICC range categories are: < 0.40 = poor; 0.40 – 0.59 = fair; 0.60 – 0.74 = good; and > 0.74 = Excellent[3]. Values above about 0.7-0.8 are considered acceptable for applied tests.

1. Nunnally, J. C. Psychometric theory (2nd ed). 1978, New York: McGraw-Hill.
2. McGraw, K. O., & Wong, S. P. Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1996:1(1), 30-46.

3. Cicchetti D.V., and Sparrow, S.S. Developing criteria for establishing the interrater reliability of specific items in a given inventory. American Journal of Mental Deficiency, 1981:86, 127-137.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 2a2.3a. Assessment of Family Psychosocial Well-Being (AFPWB): Content, Scoring and Internal Consistency, Online PHDS, all providers and top 5 individual providers with highest number of surveys

What is measured	Scoring	Internal Consistency (Cronbach's Alpha)					
Two multi-part items assess the degree to which care is provided in a family centered manner. Includes respect, understanding specific needs of child and concerns of parent, asking how feeling as a parent, understand family "culture" and talking about resources and issues in the community	Mean score on a multi-item scale	All providers	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
			1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
		0.88*	0.94*	0.75*	0.93*	0.95*	0.90*

*Met criteria for reliability and internal consistency.

Table 2a2.3b. Assessment of Family Psychosocial Well-Being: Content, Scoring and Internal Consistency, KPWN study, all providers and top 5 individual providers with highest number of surveys

What is measured	Scoring	Internal Consistency (Cronbach's Alpha)					
This measure assesses whether providers asked parents (1) if they feel depressed, sad or have crying spells; (2) whether they have someone they can turn to for emotional support; and (3) if there have been any recent changes or stressors for the parent or family (see AFPS Data Dictionary).	Mean score on a multi-item scale	All providers	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
			7 (n=80)	53 (n=77)	4 (n=74)	1 (n=67)	43 (n=66)
		0.81*	0.84*	0.82*	0.86*	0.76*	0.72*

*Met criteria for reliability and internal consistency.

Chronbach's alpha for the Family Psychosocial Well-Being measure is 0.88, ranging between 0.75-0.95 across providers with highest number of surveys (online PHDS). These findings are consistent with the findings of the KPNW study and previous peer-reviewed publications.[4,5] Inter-unit reliability coefficient for the measure scale is in the borderline of the recommended threshold (0.70) to be able to detect reliable differences between providers. Intraclass correlation coefficient for the measure is 0.76, indicating that 75.6% of the variance in the mean of the providers is "true" rather than due to chance.

4. Bethell C, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. Pediatrics. 2001 May;107(5):1084-94.
5. Christina Bethell, PhD, MPH, MBA; Colleen H. Peck Reuland, MS; Neal Halfon, MD, MPH; Edward L. Schor, Measuring the Quality of Preventive and Developmental Services for Young Children: National Estimates and Patterns of Clinicians' Performance. Pediatrics, 2004, 113(6):1973-83

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Psychometric analyses demonstrated that the Assessment of Family Psychosocial Well-Being quality measure has strong internal consistency (Cronbach's alpha ranged 0.72-0.95 across individual providers and two data sources). The quality measure provides a reliable assessment of the provision of nationally recommended well-child care with strong inter-

unit reliability coefficient (0.70) and intraclass correlation (0.76). Two different data sources indicate the AFPS quality measure provides psychometrically reliable assessment of the provision of nationally recommended well-child care with strong internal consistency.

2b2. VALIDITY TESTING

2b2.1. What level of validity testing was conducted? (*may be one or both levels*)

☒ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

A standard, multistage process was used to ensure validity of the AFPS measure:

- Focus groups and in-depth cognitive interviews were conducted throughout the survey development process;
- A review of literature identified through Medline or during key informant interviews; and,
- Three Advisory Groups comprised of pediatricians, family practitioners, consumer representatives, public health experts, and researchers, regularly reviewed and provided input on the identification of quality measurement topics and the development of the PHDS.

A "gold standard" does not exist for determining the criterion validity of patient-reported measures of quality. However, to ensure the validity of the AFPS quality measure results, we followed rigorous procedures representing best practices within the field to develop the survey questions. To ensure the content validity of measures of parent experiences, we used qualitative methods, including both focus groups and cognitive interviews, to inform development and evaluation of the Assessment of Family Psychosocial Well-Being items and composite measure.

Focus groups with families aimed to identify the aspects of health care quality that are important to parents in the area of preventive care for their children. In-depth cognitive testing of the draft survey items was conducted with 15 families representing a range of racial, income and education groups as well as different types of health insurance coverage, age of child, age and sex of parent, and number of children in family. Focus groups and cognitive interviews with 35 health care providers in Vermont and Washington and 20 parents of young children in Vermont were conducted to inform item-reduction, administration specifications, and reporting templates. Survey modifications were made based on findings in order to improve the reliability, validity and cognitive ease of the Assessment of Family Psychosocial Well-Being measure items.

Factor analysis was conducted to assess the construct validity of the AFPS quality measure. A Scree test was used to determine the number of factors to extract. Both oblique and orthogonal rotations were evaluated with promax and varimax methods used, respectively.[1] Acceptable level of factor loading for instruments developed for research purposes can be as low as 0.60[2] and factor loading more than this threshhold is considered as a strong association.[3]

To assess the concurrent validity of the AFPS quality measure, hypothesized associations among PHDS items and scales were examined using logistic regression model (KPNW Study data). We tested the following hypotheses:

- Respondents who indicate that providers talked with them about recommended anticipatory guidance topics more likely to report that the provider discussed about family psychosocial issues compared to those who did not talk with provider about the anticipatory guidance topics

- Respondents who received family-centered care more likely to report that the provider discussed about family psychosocial issues than who did not receive family-centered care.

Pearson correlation coefficients were calculated across all scale measures to test hypotheses about expected relationships among the PHDS quality measures and to assess the degree to which each of the PHDS quality measures provide unique information. We expect a moderate or strong correlation between the family psychosocial assessment scale measure (>0.30) and inter-scale correlation coefficients to be less than 0.80.

1. Bethell C, Peck C, Schor E. Assessing health system provision of well-child care: The Promoting Healthy Development Survey. Pediatrics. 2001 May;107(5):1084-94.
2. Suhr D and Shay M. Guidelines for reliability, confirmatory and exploratory factor analysis. Accessed at: http://www.wuss.org/proceedings09/09WUSSProceedings/papers/anl/ANL-SuhrShay.pdf. Retrieved 02/01/2017
3. Costello A.B and Osborne J.W. Best Practices in Exploratory Factor Analysis: Four recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation. 2005:10(7). Accessed at: http://www.pareonline.net/pdf/v10n7.pdf, Retrieved 02/01/2017

2b2.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Using behavior coding methods, for each item in the AFPS quality measure, instances where the respondent required clarification or did not appropriately answer an item were noted. Also, items where the interviewer had difficulty asking the question without edits to the wording were noted. Data analysis was used to inform item-reduction. Content was revised and refined iteratively with each set of interviews.

Cognitive testing confirmed the readability of the AFPS for people across a range of educational levels. Parents were uniformly able to complete the self-administered survey in 10-15 minutes. Readability assessments indicated the items to be written at the 8th-9th grade reading level. Survey design and formatting was finalized with input from a group of experts and family representatives.

Each of the items used to construct the AFPS quality measure was used in the factor analysis. The lowest average factor loading of 0.49 for the measure is attributable to uniformly low scores with little variation observed across individuals or health plans for the survey items comprising the family assessment quality measure. Factor analysis suggests that the scale items are unidimensional.

Logistic regression analysis showed that parents who reported that their questions on specific anticipatory guidance topics were answered were more likely to report that the provider discussed about family psychosocial issues compared with parents answering “no, but I wish we had discussed that” (odds ratio [OR]: 2.08, 95% confidence interval [CI]: 1.56-2.79, p<0.001). Moreover, parents who received family-centered care more likely to report that the provider assessed the family psychosocial issues (odds ratio [OR]: 1.71, 95% confidence interval [CI]: 1.11-2.63, p<0.02) compared to who did not receive family-centered care. In addition, findings were similar to other studies that parents report that they are rarely asked about psychosocial issues, including gun safety or how parenting works into their daily activities.[4,5,6]

Correlations between the AFPS and other PHDS quality measures were not so high as to suggest redundancy across measures (average correlation: 0.34). As expected, the highest correlation observed was between the “Assessment of family psychosocial well-being” & “Assessment of smoking, drug and alcohol use and safety in the family” (0.54) and “anticipatory guidance from providers” & the “family-centered care” measures (0.52).

Table 2b2.3. Pearson Correlation Coefficients among PHDS Quality Measures (online PHDS)

Scale Measures	Anticipatory Guidance	Family Centered Care	Ask About Parental Concern	Assessment of smoking, drug and alcohol	Assessment of family

	and Parent Education			use and safety in the family	psychosocial well-being
Family Centered Care	.52				
Ask About Parental Concern	.16	.14			
Assessment of smoking, drug and alcohol use and safety in the family	.16	.13	.07		
Assessment of family psychosocial well-being	.19	.16	.09	.54	

Average correlation: 0.34

Some of the AFPS items have been used in two national surveys of parents—The National Survey on Early Childhood Health and the National Survey of Children's Health. The PHDS-derived quality measures are among the few recognized in the Agency for Healthcare Research and Quality's Child Health Toolbox and the National Quality Measures Clearinghouse as measures that meet basic criteria for use as standardized indicators of health care quality for children.

4. Young KT, David K, Schoen C. Listening to parents. A national survey of parents with young children. Arch Pediatr Adolesc Med. 1998;152: 255–262
5. Kahn RS, Wise PH, Finkelstein JA, et al. The scope of unmet maternal health needs in pediatric settings. Pediatrics. 1999;103:576 –581
6. Kemper KJ, Osborn LM, Hansen DF, Pascoe JM. Family Psychosocial screening: Should we focus on high-risk settings? J Dev Behav Pediatr. 1994;15:336 –341

2b2.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The AFPS quality measure provides conceptually and psychometrically valid assessment of the provision of nationally recommended preventive care services for young children. Each of the five composite quality measures provides unique information about performance. Regardless of the population group or the aspect of health care assessed, the quality of health care rarely can be represented accurately by either a single composite performance measure or by assessing whether a single recommended service is provided. Components of the AFPS quality measure are used in national surveys and recognized as questions that meet basic criteria for use as standardized indicators of health care quality for children. The measure serves as an important complement to existing quality measures.

2b3. EXCLUSIONS ANALYSIS

NA ☒ no exclusions — *skip to section 2b4*

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

Not applicable

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

Not applicable

2b3.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis.*

*Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

Not applicable

2b4. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b5.

2b4.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with** Click here to enter number of factors **risk factors**

☒ **Stratification by** variable number of **risk categories**

☐ **Other,**

2b4.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

Not applicable

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

The AFPS quality measure does not require risk adjustment because we do not expect variation in the quality of care provided for children due to risk factors, e.g. children with special health care needs. The performance should be the same regardless of risk factors. The national experts extensively reviewed the risk adjustment requirements during development of the AFPS items and did not recommend risk-adjustment for the AFPS measure. In addition, during the KPNW study, we did assessment of whether the probability of receiving guidance, education or screening was higher according to a child's level of need or risk, thereby indicating that providers are customizing care to children. The study found no evidence that providers customize care to children most at risk.

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*)

Identification of variation in quality measures across subgroups of children helps to highlight aspects of care and population of children for which preventive and developmental services may be most need of improvement. Although no stratification is required (number of surveys for each individual providers may not be sufficient to stratify), the Promoting Healthy Development Survey (PHDS) includes a number of variables that allow for stratification of the quality measures by possible vulnerability:

- Child demographic characteristics (e.g., the child's age, race)
- Child health and descriptive characteristics (e.g., children at high risk for developmental, behavioral or social delays, special health care needs)
- Parent health characteristics (e.g., children whose parents are experiencing symptoms of depression)

Based on extensive literature review and expert panel, we identified that child and parent demographics such as age, sex, race-ethnicity, income, insurance, parent behavior, CSHCN screener and follow-up for children at risk can be used for stratification. Several studies have documented differences in access and quality of care provided to children, as well as in parent-reported satisfaction with care.[1-2] One study found: "Overall, 94.0% of parents reported 1 or more unmet needs for parenting guidance, education, and screening by pediatric clinician(s) in 1 or more of the content of care areas evaluated (including assessment of family psychosocial screening). Uninsured children and children aged 18 to 35

months are disproportionately represented among the 15.3% of children whose parents indicated an unmet need in each of the 4 areas of care. There are significant variations in performance on the basis of child age, race, insurance status, maternal education, marital status, and parent language as well as other factors."[3]

The KPNW study assessed child and family characteristics to characterize the child and their family based on the PHDS item responses: child's race/ethnicity, birth order, risk for developmental, behavioral, or social delays using responses to Frances Glascoe's Parents' Evaluation of Developmental Status (PEDS) items included in the ProPHDS 29 parent's education; and whether he/she is experiencing symptoms of depression using Kathy Kemper's screening items. Adjusted odds ratios were calculated using logistic regression analysis in order to assess differences in the odds of meeting quality measure criteria according to child, family and provider characteristics, after controlling for other variables.

References:
1. Halfon N, Regalado M, Sareen H,Inkelas M, Reuland CH, Glascoe FP, Olson LM. Assessing development in the pediatric office. Pediatrics. 2004 Jun;113(6 Suppl):1926-33.
2. Weech-Maldonado R, Morales LS, Spritzer K, Elliott M, Hays RD. Racial and ethnic differences in parents' assessments of pediatric care in Medicaid managed care. Health Serv Res. 2001 Jul;36(3):575-94.
3. Bethell C, Reuland CH, Halfon N, Schor EL. Measuring the quality of preventive and developmental services for young children: national estimates and patterns of clinicians' performance. Pediatrics. 2004 Jun;113(6 Suppl):1973-83.

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Not applicable

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS factors (e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects)

See 2b4.3.

2b4.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Pearson's chi-square test was used to compare the prevalence of the AFPS quality measure across the stratification characteristics. We preformed logistic regression analysis in order to assess differences in the odds of meeting quality measure criteria according to child, family and provider characteristics, after controlling for other variables.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**
Not applicable

2b4.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):
Not applicable

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:
Not applicable

2b4.9. Results of Risk Stratification Analysis:

Online PHDS: Variation is observed according to a child's age; race/ethnicity; level of risk for developmental, behavioral, or social delays across and respondent education level. Non-Hispanic white children are less likely to meet criteria on the Family Assessment measures. Children of lower educated mothers and children at high risk for developmental delay are more likely to have high Family Assessment scores.

Table 2b4.9a. Assessment of family psychosocial well-being (asked about all items) by child demographic and other characteristics

Characteristics	All children	
	n	%
Age groups		
3-8 months	779	34.7%
9-18 months	432	22.9%
19-48 months	279	20.1%
p values (Pearson chi-square)	-	<0.0001
Gender		
Male	188	30.3%
Female	191	30.5%
p values (Pearson chi-square)	-	0.96
Race/ethnicity		
Hispanic	610	31.3%
White non-Hispanic	664	23.3%
Black non-Hispanic	33	33.0%
Asian non-Hispanic	30	27.3%
Other/Multi race, non-Hispanic	24	28.2%
p values (Pearson chi-square)	-	<0.0001
Adult survey responds education level		
Did not complete high school	227	37.1%
Completed high school or higher education	1213	25.5%
p values (Pearson chi-square)	-	<0.0001
CSHCN status		
Non-CSHCN	1333	26.9%
CSHCN	157	27.6%
p values (Pearson chi-square)	-	0.71
At risk for developmental delay (online only)		
Low/No risk	476	21.3%
High/Moderate risk	216	29.2%
p values (Pearson chi-square)	-	<0.0001

KPNW study: After controlling for other child and family demographic and health factors and provider characteristics, the likelihood (or adjusted odds ratio-AOR) that a child met quality measure criteria differed significantly according to: (1) child's age and (3) child's birth

Table 2b4.9b: Mean number of developmental services care components for which AFPS quality care was received and the proportion of children meeting criteria for receiving quality developmental services by characteristics of children and families.

Characteristic of Child or Child's Family	% Meeting AFPS Criteria
Child's Age	
Less than 9 mos.	56.4% [S]
10 to 18 mos.	37.1% AOR: .47
19 to 49 mos.	30.8% AOR: .35

Child's Gender	
Male Child	37.8% [NS]
Female Child	38.5%
Child's Race	
White, Non-Hispanic	38.5% [NS]
Asian, Non-Hispanic	42.4%
Hispanic	37.8%
Other Race, Multiple Race	36.9%
Birth Order	
Not First Born	34.9% [S]
First Born	41.7% AOR: 1.29
Child's Risk for Developmental, Behavioral or Social Delays (Using Parent's Evaluation of Developmental Status)	
Low/No Risk	37.1% [NS]
At Risk	39.7%
Respondent Education	
More than High School	38.0% [NS]
High School or Less	40.9%
Respondent's Risk for Depression (Using the Kemper Screener)	
No Symptoms of Depression	37.8% [NS]
Symptoms of Depression	41.3%

NOTE: Adjusted odds ratios (AOR) derived from regression analyses listed in the table are shown only if they are statistically significant. AOR uses the first subgroup of each characteristic as a reference.

[S] differences significant at the $p < .05$ level of significance.

[NS] differences not significant.

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

The demographic and socioeconomic survey items included in the AFPS quality measure make it possible for providers to identify populations and subgroups for which health service delivery improvement is most needed.

2b4.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)
Not applicable

_____

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

Differences in the AFPS quality measure across providers is demonstrated for (1) 5 top individual providers with the highest number of surveys using Online PHDS data; (2) across 56 providers using KPNW study data; and (3) pre-post changes across time (2010-2012) after small intervention using HRSA study data for illustrative purpose.

Online PHDS: The performance scale for the AFPS quality measure was calculated using the scoring methods described in Attachment A-4. Individual provider level differences in performance were illustrated by the proportion of children meeting the AFPS quality of care criteria across 5 top providers with the highest number of completed surveys after their well-child visit.

KPNW Study: The significance of differences observed in the proportion of children meeting criteria for the quality measure across pediatric providers (n=56) was evaluated using t-tests. The relative spread in the AFPS score across providers was assessed using the coefficient of variation statistics (standard deviation across providers multiplied by 100%). Multi-level regression models were conducted using the pediatric provider as the level 2 clustering variable, in order to assess the degree to which the probability that a child meets criteria on each quality measure is explained by differences between providers (called the "clustering effect"). In implementing this multi-level regression method (Empty Model), the presence of a significant clustering effect by pediatric providers was estimated prior to accounting for the child and family characteristics associated with each provider. Second, variables related to the child and family characteristics (child's age, gender, race/ethnicity, birth order, developmental and behavioral delay risk status; parent education and risk for depression) were added to the Empty Model to assess how much of the provider clustering effect observed remains after accounting for these characteristics (called the "Patient Model").

HRSA study: Quantitative data results for the baseline (2010) and follow-up (2011-12) study of the intervention sites using the HRSA Evaluation Study data were conducted using basic descriptive statistics to describe each sample and applying chi-square test of statistical significance to assess differences in the quality measure for the baseline and follow-up samples.

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Online PHDS:
Table 2b.5a presents the proportion of children whose parents were screened for family psychosocial issues across the 5 providers with the highest number of completed PHDS surveys. Variation across providers who asked parents about all

survey items related to family emotional and mental health issues is substantially wide across observed 5 providers. Only 6.6%-44.3% of parents of young children reported that their child's pediatric clinician discussed psychosocial topics such as parent emotional well-being and partner support in parenting.

Table 2b5.2a: Proportion of Parents Screened for Family Psychosocial Issues Meeting AFPS Measure Criteria: Top 5 individual providers with highest number of surveys

Characteristics	All Children (n=5355)	Provider IDs for 5 individual providers with highest number of surveys (number of surveys)				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
Asking parents about at least one of three emotional and mental health issues (at least one of these topics)	53.0%	39.1%	25.3%	51.8%	54.4%	47.3%
Asking parents about at least one of three emotional and mental health issues (asked all items)	27.0%	28.3%	6.6%	37.3%	44.3%	29.7%

KPNW Study: 38.1% reported that providers discussed at least one emotional and mental health topics. Range across providers in the proportion of children who met quality measure criteria was substantial and statistically significant (p=0.002)

Table 2b5.2b: Proportion of all children in the study who met criteria for receiving quality services and ranges in proportion across providers. (SD=Standard Deviation)

Developmental Services Quality Measures	Proportion of All Children Meeting Measure Criteria (n = 2173)	Range in the Proportion of Children Meeting Measure Criteria Across 51 Pediatric Providers	Relative Variation (COV) in Measure Scores Across Pediatric Providers
Asking parents about at least one of three emotional and mental health issues (at least one of these topics)	38.1%	18.5% to 65.2% SD: 10%; (p = 0.002)	26.1%

Only providers with 15 or more completed PHDS surveys are included in the provider level analysis. The number of completed surveys across providers ranged from 15 to 153.

Multi-level analysis: For the Empty Model that used the provider as the level 2 clustering variable, only 1.1% to 2.2% of the total variance observed in whether children met criteria for each quality measure was explained by either measured or unmeasured differences between the providers that they see. This suggests that there is nearly as much variation across children seeing the same provider as across children seeing different providers. These findings translate into a 1.19 to 1.29 median odds ratio across the six quality measures, indicating that the odds of meeting quality measure criteria if the average child were to transfer from the lowest to the highest performing provider is 1.19 to 1.29 across the all quality measures. When child/family level characteristics are added to the model (Patient Model), the total variance explained by differences between providers does not change significantly.

HRSA study

The HRSA study found statistically significant and positive changes for the study interventions (providers attended a training session on Bright Futures guidelines at the meeting) based on the AFPS quality measure. Parents were more likely to be asked about one or more psychosocial (family assessment) topics, including mental health and emotional support, at follow-up. The tables below present comparison of percent of children who received care met the quality care criteria between baseline and follow-up survey data.

Table 2b5.2c. AFPS by Children's Characteristics: ---Parent was asked about one or more Family Psychosocial Topics

Characteristics	Baseline % (n)	Follow-up % (n)	Chi-square test

			p value
Age			
3-9 months	23.9% (132)	45.0% (145)	<0.0001
10-18 months	21.5% (87)	34.4% (111)	<0.0001
19-48 months	29.3% (103)	50.5% (112)	<0.0001
Race			
Hispanic	26.0% (26)	47.8% (22)	0.01
White	24.5% (252)	42.7% (305)	<0.0001
Asian	28.6% (8)	(4)	-
Multiple or other	50.0% (12)	50.0% (9)	1.00
Insurance type			
Private or private and public	22.9% (248)	37.9% (257)	<0.0001
Public only (includes Medicaid, Medicare, CHIP, and Military)	33.5% (64)	39.1% (63)	<0.0001
Other insurance type	(2)	(1)	N/A
Uninsured	(4)	(3)	N/A
At risk of developmental delay			
Low/no risk	24.3% (248)	41.6% (227)	<0.0001
High/moderate risk	26.0% (73)	44.7% (76)	<0.0001

2b5.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

Significant gaps and unexplained variations remain in the quality of developmental services for young children. The probability of receiving AFPS varies nearly as much across children seeing the same provider as across providers. The quality measure assessed here provide a relatively comprehensive picture of performance in the area of preventive and developmental services for young children.

2b6. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without SDS factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without SDS factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b6.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

Not applicable.

2b6.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)
Table 2: Proportion of all children in the study who met criteria for receiving quality developmental services across six components of care and ranges in proportion across providers and offices. (SD=Standard Deviation)

Not applicable.

2b6.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

Not applicable.

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

The AFPS items were developed based on several rounds of cognitive interviews with parents to ensure quality of responses appropriate to the questions and minimize missing responses.

Online PHDS: Rate of survey completion was calculated based on survey start and complete dates for each respondent. According to the quality measure scoring protocol, if a parent answered less than half of the items in the quality measure, their score is considered to be missing. This does not include items that should have been appropriately skipped. Missing responses are not given a valid score and are not included in the calculation of the quality measure.

KPNW Study: Of the 5,755 sampled children, 2,173 surveys were returned (37.8%). For these children, the provider the parent identified and the provider to which the child was assigned by the health plan were the same 97.3% of the time. A 95% response rate was obtained for the provider survey.

2b7.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Online PHDS data show that 2.6% of parents who started the survey did not complete the survey.
Table 2b7.2a presents the frequency of missing values for the Assessment of Family Psychosocial Well-Being measure.

Table 2b7.2a. The frequency of missing values for Assessment of Family Psychosocial Well-Being measure, overall and top 5 providers

Quality measures	Overall	Provider ID				
		1029 (n=94)	948 (n=91)	1067 (n=90)	927 (n=79)	1030 (n=77)
	% (n)	% (n)	% (n)	% (n)	% (n)	% (n)
Assessment of Family Psychosocial Well-Being	2.6% (149)	2.1 (2)	0	7.8 (7)	0	3.9 (3)

KPNW study: Children whose parents responded were not different from those who did not respond in terms of their gender and insurance type. The responding population were slightly less likely to be in the 19 to 48 month age group (55.8% sampled, 51.5% responding) and were somewhat more likely to have had more than one well-visit in the past (67.5% sampled, 74.7% responding).

Table 2b7.2b. Sociodemographic Characteristics of KPNW Starting and Responding Sample

Characteristic	Proportion of Starting Sample (N=5755)	Proportion Respondents as of (N=2162)
Gender of Child[NS]		
Male child	52.7	53.7
Female child	47.3	46.3
Age of the Child[S]		
Child age 3-9 months	19.4	21.8
Child age 10-18 months	24.9	26.7
Child age 19-48 months	55.8	51.5
Type of Insurance[NS]		
Private	98.6	98.5
Public	1.4	1.5
Child's Health Care Utilization		
Number of well-child visits [S]		
1 Well-Child Visit	32.5	25.3
2 or More Well-Child Visits	67.5	74.7
Number of emergency room/urgent care visits		
0 ER/urgent care visits	49.8	51.0
1 ER/urgent care visit	26.2	25.8
2 or more ER/urgent care visit	24.0	23.2
Number of overnight hospital stays [NS]		
0 overnight hospital stays	96.6	96.9
1 or more overnight hospital stays	3.4	3.1

[S]Denotes variables for which statistically significant variation exists between the starting and responding sample for the target child or respondent characteristic.

[NS]No significant variation exists between the starting and responding sample for the target child or respondent characteristic.

2b7.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Information about non-respondent is not available to compare with those who responded the survey because online PHDS is publicly available tool. However, the low rate of incomplete survey (2.6%) suggests that the measure was acceptable to respondents. Overall, the quality measure had 2.6% of missing cases, ranging 0-7.8% across the top 5 providers with the highest number of surveys. Few overall missing values suggest that the measure level results unlikely to be biased by non-response to the survey questions.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. **Data Elements Generated as Byproduct of Care Processes.**

Other

If other: Data are generated by parents completing the CAHMI developed Promoting Healthy Development Survey (PHDS), which is sent to them by their provider

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. **To what extent are the specified data elements available electronically in defined fields** (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

Patient/family reported information (may be electronic or paper)

3b.2. **If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources.** For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

Data are based on parent experience and thus require a parent-report mechanism. Parents use the CAHMI-developed Promoting Healthy Development Survey (PHDS). CAHMI captures the data at the provider level through a process described above and in the Evidence Form, Figure 1.

3b.3. **If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.**

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. **Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.**

IF a PRO-PM, consider implications for both individuals providing PRO data (patients, service recipients, respondents) and those whose performance is being measured.

During 2012-2016 we have experienced some operational delays. in the past several years. In 2012, the provider feedback reports were not automated. When providers wanted a summary report, CAHMI had to manually create them. This was excessively time consuming and CAHMI did not have resources to continue the manual generation of the reports. We sought and received funding to automate the reports. Some difficulties with contractors and staff change-over caused major delays in the project. Then, CAHMI moved from the Oregon Health & Sciences University to Johns Hopkins University School of Public Health in 2014, and it was necessary to upgrade the CAHMI servers. No technical support was available for the transition which caused further delays. Additionally, the PHDS was originally developed in 2001; thus much of the coding and back-end technology for this tool was antiquated and ceased to function after the move. Consequently, and as a result of new improved technology, we have had to redesign the two PHDS related websites - the PHDS toolkit and the parent survey -- as well as the CAHMI PHDS database. Lack of funding caused delays. However, we anticipate launching the new PHDS in February 2017.

Time and cost of data collection are low: provider registration takes about 10 minutes and the parent survey takes about 15-20 minutes to complete. To date, implementation has been limited by lack of funding and resources for outreach, communication and technical support. Our experience in the development and evaluation of the PHDS demonstrated a clear and compelling need to work closely with providers to overcome the many myths that both parents and providers have about patient-engagement quality improvement tools. For the PHDS to be adopted by providers, it is essential to demonstrate, for example, that tool adds value for both the parent and provider, that it fits into and typically improves work flow in the office; improves parent-provider communication, and most important, improve the quality and delivery of nationally recommended services for children. This can only really be accomplished by collaboration and partnership with providers.

3c.2. **Describe any fees, licensing, or other requirements to use any aspect of the measure as specified** *(e.g., value/code set, risk model, programming code, algorithm)*.
None

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. **Current and Planned Use**

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Professional Certification or Recognition Program Quality Improvement (Internal to the specific organization)	

4a.1. **For each CURRENT use, checked above (update for maintenance of endorsement), provide:**

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

NA

4a.2. **If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons?** (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
The PHDS toolkit (www.phstoolkit.org) and the parent-reported PHDS (www.wellvisitsurvey.org) were used by 68 uniquely identified providers across the country through 2013. We are happy to provide a list of these providers to NQF if desired. In 2014, CAHMI moved from the Oregon Health & Sciences University, Portland OR to the Johns Hopkins University, Baltimore, MD. As a result of the move, and because both server and database technologies had rapidly evolved and improved over the past few years, it was necessary to upgrade our servers, which in turn caused some technical issues with the links between the provider toolkit, the PHDS, and the CAHMI PHDS database. Additionally, the PHDS was originally used to compare providers within a practice as well as between practices within a health system. The anticipated use of the Online PHDS is intended to provide feedback only for individual providers and at the clinic or practice level but not between providers. The combination of these factors led to a decision to upgrade and redesign the PHDS toolkit, PHDS database and Parent Survey. (The PHDS parent survey itself, however, remains fully operational, although use has been nominal from 2014-present, and can be accessed at

www.wellvisitsurvey.org.) The redesign required additional time, IT and CAHMI staff resources and delays were incurred during 2014-2015. However, we are now in the process of finalizing the PHDS Toolkit and database redesign, which is anticipated to be completed and launched in February 2017.

4a.3. **If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement.** (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

The redesigned PHDS system (registration toolkit, parent survey tool and PHDS database) is anticipated to be completed and fully functional by February 2017. We have a communication and outreach plan to promote the PHDS as part of the CAHMI Cycle of Engagement (see Attachment A-1: Item #4), which includes the CAHMI Well Visit Planner (www.wellvisitplanner.org) -- a free parent engagement tool that helps prepare parents for the upcoming well child visit – and the post-visit PHDS which assesses whether the parent received services in alignment with national guidelines as well as family centered care. We have been promoting the Cycle of Engagement in national meetings (AMCHP, PAS, APHA, AcademyHealth ARM, National Child Heath Policy Meeting, and more) over the past several years. We presented the Cycle of Engagement at the CMS Quality Meeting December 13, 2016 and have further plans to unveil the redesigned version at meetings in 2017. The WVP and PHDS have also been endorsed tools that meet requirements for Bright Futures implementation.

We have received substantial interest in the CAHMI parent-engagement tools (both the WVP and the PHDS) from and are in extensive conversations with a number of organizations and agencies including health systems, payers, provider organizations – (CMS/Medicaid, Title V, Head Start, Kaiser Permanente and others); professional associations such as the American Academy of Pediatrics, Bright Futures, National Medicaid Medical Directors, the Children's Hospital Association (CHA), AcademyHealth, Association of Maternal and Child Health Programs (AMCHP), CityMatCH, National Initiative for Children's Healthcare Quality (NICHQ), Autism Speaks, Prevent Child Abuse America; National Prevention Information Network (NIPN); national community-based programs and organizations; philanthropic funders; software platform and electronic medical records systems developers and family organizations. We are in the process of securing funding for Cycle of Engagement EMR integration and implementation projects in partnership with or from a number of interested parties. Further, we are finalizing our application to the American Board of Pediatrics to have the Online PHDS certified as a web-based Maintenance of Certification (MOC) (Part 4) quality improvement (QI) tool for pediatricians. ABP has expressed significant interest in the PHDS and provided some initial funding for the redesign efforts.

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b. **Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)**

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

In a 2010-2012 study of a large pediatric practice in Oregon (n=551 providers), family assessment increased from 21.5% (n=116) at baseline (2010) to 43.7% (n=111, 2011-12, AOR=3.32, 2.24-4.91) post implementation of the CAHMI Well Visit Planner - a family engagement tool to assist parents in planning for their well child visit. This represents a 103.3% increase and was statistically significant at the 95% confidence level. The PHDS, which contains Assessment of Family Psychosocial Screening measure, was used as the evaluation tool.

4c. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4c.1. **Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.**

There were no unintended or unexpected consequences that we are aware of.

4c.2. **Please explain any unexpected benefits from implementation of this measure.**
There were no unexpected benefits that we are aware of.

4d1.1. **Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.**
How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
Extensive qualitative interviews with providers and parents have been conducted and previously reported (See Attachment 2, Evidence Report)

4d1.2. **Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.**
Key informant interviews and focus groups with parents and providers were held throughout the testing and evaluation period. We obtained baseline and post-implementation information from providers and post-implementation information from parents. It was necessary to work closely with practices to demonstrate value of the family engagement tools (Well Visit Planner and PHDS) as well as to modify the process to fit individual practice office culture and work flow. A significant amount of provider and staff education was needed to overcome fears and myths that the tool would add to, not help, time management and that parents would not want to participate. This was accomplished by continued and persistent relationship building, spending much time in the office setting with the staff and providers and holding frequent Q&A sessions as the process unfolded.

4d2.1. **Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.**
Describe how feedback was obtained.
Through key informant interviews and focus groups with parents and providers.

4d2.2. **Summarize the feedback obtained from those being measured.**
The Assessment of Family Psychosocial Screening measure is seen by providers as an excellent way by which they can improve the quality of the well child visit. In particular this matters a great deal to the providers who are being financially incentivized for family-centered care outcomes.

4d2.3. **Summarize the feedback obtained from other users**
For the most part, parents appreciated being asked about their experience with their well child visits and used it as a way to provide confidential feedback to the providers.

4d.3. **Describe how the feedback described in 4d.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.**
The feedback was helpful for future implementation efforts of CAHMI's family engagement tools. The feedback, however, did not result in any changes to the measure itself.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures
Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.
No

5.1a. **List of related or competing measures (selected from NQF-endorsed measures)**

5.1b. **If related or competing measures are not NQF endorsed please indicate measure title and steward.**

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. **If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):**

Are the measure specifications harmonized to the extent possible?

Yes

5a.2. **If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.**

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. **If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):**

Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

NQF #0011 - the PHDS (Promoting Healthy Development Survey) - was endorsed by NQF on October 4, 2012. The PHDS contains the AFPS measure. Neither the questions nor the scoring of the questions have changed since the PHDS was endorsed. It is not actually a competing measure; rather, the AFPS measure is embedded in the PHDS tool.

Please note: The PHDS endorsement (#0011) can be found on the NQF measures website but does not appear to be found in the NQF directory in Question 5 above. Hence, we were forced to enter a "no" to Q5 in order to submit this application.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment **Attachment:** Attachment_A_Supplemental_Materials_Revised_01_18_17.pdf

Contact Information

Co.1 **Measure Steward (Intellectual Property Owner):** Child and Adolescent Health Measurement Initiative

Co.2 **Point of Contact:** Christina, Bethell, cbethell@cahmi.org, 443-287-5092-

Co.3 **Measure Developer if different from Measure Steward:** Child and Adolescent Health Measurement Initiative

Co.4 **Point of Contact:** Christina, Bethell, cbethell@cahmi.org, 443-287-5092-

Additional Information

Ad.1 **Workgroup/Expert Panel involved in measure development**

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

National Advisors for Patient Centered Quality Improvement of Well-Child Care:
Betsy Anderson, Family Voices
David Bergman, Stanford University
Dimitri Christakis, University of Washington
Paula Duncan, University of Vermont
Cynthia Minkovitz, Johns Hopkins School of Public Health
Amy Perritti, American Academy of Pediatrics

Ed Schor, The Commonwealth Fund
Judy Shaw, University of Vermont
Sara Slovin, Johns Hopkins Medicine

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 **Year the measure was first released:** 2002

Ad.3 **Month and Year of most recent revision:** 01, 2017

Ad.4 **What is your frequency for review/update of this measure?** 3 years

Ad.5 **When is the next scheduled review/update for this measure?** 01, 2018

Ad.6 **Copyright statement:** None

Ad.7 **Disclaimers:** None

Ad.8 **Additional Information/Comments:** None