NATIONAL QUALITY FORUM

MEASURE WORKSHEET

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

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

Purple text represents the responses from measure developers.

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

Brief Measure Information

NQF #: 3599

Corresponding Measures:

De.2. Measure Title: Pediatric Asthma Emergency Department Use

Co.1.1. Measure Steward: Albert Einstein College of Medicine

De.3. Brief Description of Measure: This measure estimates the rate of emergency department visits for children ages 3 – 21 who are being managed for identifiable asthma, using specified definitions. The measure is reported in visits per 100 child-years.

The rate construction of the measure makes it a more actionable measure compared to a more traditional quality measure percentage construct (e.g., percentage of patients with at least one asthma-related ED visit). The rate construction means that a plan can improve on performance either through improvement efforts targeting all patients with asthma, or through efforts targeted at high-utilizers, since all visits are counted in the numerator. For a percentage measure, efforts to address high-utilizers will be less influential on performance and potentially have no effect at all even if a high utilizer goes from 8 visits a year to 1, since in order to improve performance, a high-utilizer has to get down to zero visits.

This measure was developed under the Pediatric Quality Measurement Program, funded by the Centers for Medicare and Medicaid Services and administered by the Agency for Healthcare Research and Quality. https://www.ahrq.gov/pqmp/about/what-is-pqmp.html

1b.1. Developer Rationale: In 2009, Congress passed the Children´s Health Insurance Program Reauthorization Act (CHIPRA, Public Law 111-3), which presented an unprecedented opportunity to measure and improve health care quality and outcomes for the nation´s children, including those enrolled in Medicaid/CHIP. When CHIPRA was enacted, the Agency for Healthcare Research and Quality (AHRQ) and the Centers for Medicare & Medicaid Services (CMS) began working together to implement selected provisions of the legislation related to children´s health care quality.

The law called for the establishment of the CHIPRA Pediatric Quality Measures Program (PQMP) to improve and strengthen the "Child Core Set" of measures and develop new measures as needed.

The proposed measure 3599 Pediatric Asthma Emergency Department Use was developed and then further tested and refined under the PQMP. Here we present the rationale for the proposed measure, informed by our work in convening state-level quality improvement collaboratives (one in CA and one in VT) specifically focused on improved asthma care, and assessing the relationship between improvements in asthma care processes and performance on the proposed

pediatric ED utilization measure. The following text draws from the PQMP Toolkit for the measure for use by clinics and health plans interested in using the measure for quality improvement efforts.

_____

What is needed for QI success:

The Pediatric Asthma Emergency Department Use measure is an outcome measure based on administrative data. In that context, the toolkit user entity (state agency, health plan, healthcare organization, improvement partnership, provider group) will need to partner with practices and quality improvement coaches to evaluate systems and develop process measures to guide improvement efforts that will impact the measure. Examples of process measures can be developed from clinical guidelines such as the National Heart Lung and Blood Institute Guidelines for the Diagnosis and Management of Asthma.

It is important to recognize that most process measures will be contained in a practice's electronic health record (EHR) or a data source separate from claims data. Additionally, effective process measures should be evaluated longitudinally to assess performance over time and allow for identification of variation, either intended or unintended.

Successful improvement requires sound quality improvement science methodology, appropriate resources and ready access to reliable data. Without these components (appropriate training, infrastructure and data access), application of QI may lead to unintended consequences, such as provider frustration or QI 'fatigue'.

See below for a summary of potential strategies to support implementing quality measurement and improvement strategies in primary care settings to reduce asthma-related ED visits from the perspective of a health plan.

Summary of Strategies and Complementary Toolkit Resources

1) Goal: Understand the population and the system resources in your care delivery area
    a. Resources Required: Understand the population and the system resources in your care delivery area
    b. Health Plan contributions: Foster partnerships and determine strategic alignment(s)
2) Goal: Partner with practices in Health Plan network
    a. Resources Required: Practice network
    b. Health Plan contributions: Engage practices in collaborative
3) Goal: Engage practice leadership
    a. Resources Required: Practice champion
    b. Health Plan contributions: Financial alignment for clinical champion(s)
4) Goal: Develop improvement science expertise
    a. Resources Required: QI Coaching
    b. Health Plan contributions: Offer financial support for QI infrastructure
5) Goal: Determine baseline performance on NHLBI measures
    a. Resources Required: Process measures from EHR
    b. Health Plan contributions: Support practices to engage EHR vendor/ practice support to obtain data
6) Goal: Assess periodic performance/improvement over time
    a. Resources Required: Periodic data pull from EHR for process measures (by practice-based clinicians or chart auditor)
    b. Health Plan contributions: Develop practice-based incentives for improvement
7) Goal: Understand variation in performance and guide improvement efforts
    a. Resources Required: Practice level strategies

b. Health Plan contributions: NA

8) Goal: Systems Learning

a. Resources Required: Practice data of children who went to ED

b. Health Plan contributions: Health plan provides practice reports on ED utilization for clinic health plan members

QI Strategies

Overview: Approaches to Quality Improvement in Asthma Care

There are numerous factors and settings that impact the asthma emergency department (ED) measure (e.g., schools, ED, acute care, access to specialists, community, etc.), and must be considered in trying to reduce inappropriate ED use for pediatric asthma. Many factors can lead to a child with asthma receiving care in the ED such as poor asthma control, severity of symptoms, decreased access to care, and ability to enact emergency care (such as use of a rescue inhaler) among many others. When thinking of these factors and where they occur, they generally can be attributed to the patient's home and school environment, medical home, the ED or a combination (Allen, 2019). Interventions engaging the ED should be considered if there is a high rate of patients with multiple visits to the ED. In this scenario, it is important to evaluate access to care, environmental factors, ED care and the connection between the ED and the medical home.

There are three general quality improvement (QI) approaches to decrease pediatric ED visits for asthma that have a strong evidence base:

- Primary Care,
- Provider Continuing Medical Education, and
- Parental and School-Based

There is insufficient evidence to recommend a single approach, or set of interventions, over another because there are many factors that influence what will be the most effective approach for a care system. Some systems with a high degree of integration and QI capacity have chosen multiple interventions to reduce ED visits (Allen, 2019). However, most care systems will likely be best served to identify a single approach after evaluating their outcome and process measures while identifying the key drivers of performance. Assembling an interprofessional team to understand key stakeholder priorities and readiness coupled with a thorough and systematic approach to QI are essential to achieving success.

This pediatric asthma measure has potential to improve asthma care, reduce ED utilization, and promote collaboration between health plans and primary care practices. Successful utilization of the measure will necessitate interpreting data from multiple sources and business entities. Because of this, there will be practical, ethical and legal limitations relative to sharing data and how improvement efforts are implemented. While the approaches described above each have merit, the PQMP grantees charged with testing how to use the pediatric ED use measure chose to focus on the intervention area with the most evidence of success. This toolkit outlines primary care-focused interventions using an intensive educational approach and methods to develop improved systems of care.

Primary Care-Focused Approach

Most interventions that have been successful in improving asthma ED outcomes through provider-based activities have included intensive educational approaches or methods to develop improved systems of care within the primary care office setting. Harder et al. examined the effects of a one-year QI collaborative for primary care clinicians that focused on office systems strategies (e.g. asthma assessment, control and management, and patient education). Compared to control practices, the participating practices noted a substantial decrease of nearly 40 percent in asthma-related ED visit rates more than a year after the end of the collaborative (Harder, 2020). The development of a systematic primary care approach to asthma care can also improve asthma health care utilization. In a pragmatic, cluster randomized controlled trial, Yawn et al. demonstrated that the use of Asthma APGAR (Activities, Persistent, triGGers, Asthma medications,

Response to therapy) tools improved rates of asthma control and reduced asthma-related ED and urgent care visits (Yawn, 2018).

PQMP Toolkit Approach: Primary Care Collaboratives

The IMPLEMENT for Child Health initiative (IMPLEMENT) is the overall program that tested out the usability of the PQMP asthma ED measure by conducting QI initiatives in both San Francisco, California (SF Collaborative) and in Burlington, Vermont (VT Collaborative), both aimed to improve pediatric asthma care delivered in a primary care setting. The strategies described in this toolkit reflect the learnings from those two QI initiatives aimed at examining the usability of the asthma measure. In the SF Collaborative, primary care practices participated in a 12-month learning collaborative. In the VT Collaborative, practices had participated in an earlier Vermont statewide asthma learning collaborative (CHAMP Learning Collaborative, for more information see https://www.med.uvm.edu/vchip/champ) and therefore a more targeted approach was undertaken – performing a "deep dive" to examine factors that contributed to high ED rates. Staff and faculty from the University of Vermont's Vermont Child Health Improvement Program's (VCHIP) provided the QI expertise for both initiatives.

_____

In summary, the proposed pediatric asthma measure is responsive to improvements in QI process measures, as demonstrated by Harder et al. (publication in press), and improvements can be driven either by individual clinics, clinics participating in a collaborative, or health plans supporting clinics in improvement efforts. Health plans could also consider addressing the social determinants of health, as described in the Logic Model section of the Evidence attachment. Addressing the social determinants of health, while not a focus of the PQMP work, is another avenue for potential intervention to improve performance on this measure.

S.4. Numerator Statement: Number of asthma-related ED visits

S.6. Denominator Statement: 100 Child Years for children with identifiable asthma

S.8. 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, including the month being assessed.

De.1. Measure Type: Outcome

S.17. Data Source: Claims

S.20. Level of Analysis: Health Plan

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

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

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

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?

Preliminary Analysis: New Measure

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or

service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary

- This is a new claims-based outcome measure at the health plan level that estimates the rate of emergency department visits for children ages 3 – 21 who are being managed for identifiable asthma.
- Developer provides a logic model that depicts the relationship between key drivers of pediatric asthma and interventions that can result in better outcomes.
- The developer provides empirical evidence of a healthcare action that improves measure performance:
    - This measure was refined and further tested in the Pediatric Quality Measures Program.
    - As part of this work, the developer assessed the relationship between improved performance on specific asthma care processes, achieved through a state-wide quality improvement collaborative in Vermont, and decreased asthma ED visits, using the specifications for the proposed measure #3599 of pediatric asthma ED visits.
    - The developer states this study provided strong evidence that improvement in asthma-related process measures is associated with improvement in performance on the proposed measure of pediatric asthma-related ED visits.

Exception to evidence

- Not Applicable

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

Outcome (Box 1)→ Relationship between the outcome and a healthcare action demonstrated by empirical data (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 provided data for two states: California and Massachusetts. Results are listed below:

State	# of plans	# of patients	Mean	SD	Min	Max	IQR
California	103	321,072	24.4	9.4	7.6	63.5	18.3-28.9
Massachusetts	29	83,577	12.7	6.7	0	27.7	9.6-18.0

- The results suggest overall relatively high mean rate of ED use among children with identifiable asthma and moderate variability in plan performance both between states as well as between plans within states.

Disparities

- The developer provided the following data for California:
    - Gender: Pediatric ED visits/100 child-years
        - Females: 26.0
        - Males: 26.1
    - Race/Ethnicity: Pediatric ED visits/100 child-years

        - White: 23.6
        - Latinx: 24.2
        - Black: 40.6
        - Asian/Pacific Islander: 15.2
        - Other: 24.3
        - Unknown: 35.9
    - The data provided demonstrates disparities in care for black children who have a much higher rate of ED visits than children of other races/ethnicities.

Questions for the Committee:

- Is there a gap in care that warrants a national performance measure?
- Disparities information is provided. Are you aware of any additional evidence that 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)

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

- This outcomes measure captures the number of asthma related ED visits per 100 child years. Construction of the metric allows improvement broadly for all children w/**asthma** or by targeting high utilizers.
- Evidence relates to ED visits for children, adolescents and young adults with Asthma - based on 100 years of life per unit of measure (practice, system, healthplan). Studies are current
- na
- This is a new outcome measure. It utilizes health care plan electronic claims to estimate the rate of emergency department visits for children ages 3 – 21 who are being managed for identifiable asthma. This article from the December 2020 issue of Pediatrics - Statewide Asthma Learning Collaborative Participation and Asthma-Related Emergency Department Use - was provided as evidence. The article's conclusion is that participation in an asthma-focused quality improvement collaborative was associated with decreased asthma-related ED visit rates (asthma-related ED visit rate in 20 participating practices' population decreased by 5.8 per 100 child-years, compared to an increase of 1.8 per 100 child-years for control practices). The introduction to the journal article cited references from the literature but there was no systematic review of the literature included in the measure submission.
- Evidence supports the measure.
- No major concerns
- Some concerns about the measure, but it passed the SMP.
- There is strong evidence that supports the measure.
- Moderate evidence rating

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

- The developer has submitted performance data from CA and MA. There was a significant difference in performance between CA and MA. There is significant racial disparity with **African** American children disproportionately impacted.
- There is reported variation across the country (Ca vs MA) and also variation based on race/ethnicity with Black population having higher ED visits. There is a nicely detailed framework (toolbox) to implement for improvement by practice or collaboratives.
- Relatively high rates; Disparities in care of black children
- The results suggest overall relatively high mean rates of ED use among children with identifiable asthma and moderate variability in plan performance both between states as well as between plans within states. The mean # of asthma related ED visits per child diagnosed with asthma per 100 child years was 24.4 for California and 12.7 for Massachusetts. California data was analyzed by gender and race/ethnicity. The authors of this application wrote that the data provided demonstrates disparities in care for black children, who have a much higher rate of ED visits than children of other races/ethnicities.
- High rate of ED use for asthma and moderate variability suggest opportunity for improvement.
- Yes, gap exists
- Did developer consider just focusing on the Black population?
- The measure developer provided performance gaps by gender, race/ethnicity. There were no disparities in gender. There were disparities in race/ethnicity, particularly between white/Latinx and black.
- Performance gap exists

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

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

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

Reliability

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

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

Validity

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

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

Composite measures only:

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

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

Evaluators: Sam Simon; Susan White; Laurent Glance; J. Matt Austin; Daniel Deutscher; Paul Kurlansky; Joseph Hyder; Alex Sox-Harris

Methods Panel Review (Combined)

Methods Panel Evaluation Summary:

This measure was reviewed by the Scientific Methods Panel and discussed on the call. A summary of the measure and the Panel discussion is provided below.

Reliability

- Reliability testing conducted at the score level:
    - Developer tested the measure using split-sample analysis and ICC calculations for score level reliability testing in 26 health plans in Massachusetts and 101 health plans in California
        - MA health plans ICC: 0.72
        - CA health plans ICC: 0.86
    - Testing was conducted using a risk-adjusted approach in a mixed effect model
    - While it was not clear which ICC was used, SMP members agreed with this approach and considered the results to be moderate to strong.
    - One SMP member called for greater clarity regarding the definition of the denominator

Validity

- Validity testing conducted at the score level:
    - Developer tested the measure for construct validity by using predicted performance for the plan-level random effect in the risk adjustment models and then transformed that into a Z-score. Pairwise correlations were made to select HEDIS measures. Correlation results:
        - Medication Management in Asthma Compliance 50% CC: 0.12
        - Medication Management in Asthma Compliance 75% CC: 0.13
        - Child Vaccines: 0.33
        - ACE Monitoring: 0.05
        - Cervical Cancer Screening: 0.04
        - Low Back Pain Imaging: 0.05
    - Predictive validity was conducted as a secondary analysis at the clinic level in Vermont, assessing a quality innovation (QI) learning collaborative reduction in emergency department (ED) utilization through a difference in difference analysis
        - Adjusted marginal ED visit rates were superior in QI participants
            - Non participants change over time was 1.58 visits per 100 person-years
            - Participants change over time was -6.28 visits per 100 person-years
            - Difference in differences was -7.28
        - This suggests that the measure is responsive to QI initiatives related to ED utilization reduction for asthma
    - SMP identified threats to validity:
        - The exclusion of testing only in data from MA.
        - Missing data: Certain elements have high level of missingness, which may partially account for the difference in model performance between Medicaid and all payer data

        - Risk adjustment: R2 highly variable between development and validation sets. Confidence intervals and p values not given. Very few variables (6) included in the model. Method of variable selection was a priori and not well explained
        - "Given that secondary diagnosis for asthma may be unrelated to the reason for ER visit/admission, the validity of the measure as constituted has not been established."
        - "Low construct validity values."
        - "Concerns remain with the number of plans that have a relative high percentage of members who are missing social risk factor data. The social risk factors are key adjustment variables in the risk adjustment models. 20,000 of 85,000 members are in plans for which 10% of more of members are missing social risk data. And concerns remain with the risk-adjustment model. Given the possible differences in SES factors for APCD and Medicaid populations, applying a singular risk model to all populations may prove challenging/problematic."
    - The SMP also noted that secondary asthma presentation was identified as a potential confounder for the measure. The developer described the issues they found associated with including secondary asthma: it is not uncommon for rhinitis or respiratory infection to result in asthma exacerbation, but there is a challenge in identifying appropriately the trigger or the condition as the primary rationale for the patient presenting in the ER.
        - The developer noted analyses revealed that fever, influenza, and upper respiratory infection are often listed as first diagnoses and asthma is second. The developer also clarified that the measure includes a "second" diagnosis of asthma and not "secondary" diagnoses. The sample size reduction was close to 50 percent ,if a second diagnosis of asthma was excluded with sensitivity analysis showing similar results. Therefore, the developer argued that inclusion of the second diagnosis is important to the measure.
        - One SMP member suggested that the positioning of the diagnosis is irrelevant and may occur beyond the second diagnosis in the coding, which may be a confounder for the measure.
        - However, the developer noted that research has shown that pediatric patients do not tend to have a lot of diagnoses and so the likelihood that asthma would appear lower down in a long list of diagnoses is unlikely.
        - The SMP encouraged developers to cite research like this within their submissions.
    - The SMP was not able to come to consensus on this measure.

Questions for the Committee regarding reliability:

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

Questions for the Committee regarding validity:

- The Scientific Methods Panel was not able to come to consensus on this measure. Do you have any concerns about the validity of the measure?

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

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

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

- reliability was acceptable.
- It appears that reliability is good, based on provided testing, however, reviewers note that there is lack of specifics of methodology (denominator data)
- na
- The developer tested the measure using split-sample analysis to assess signal to noise analysis ICC calculations for score level reliability testing in 26 health plans in Massachusetts and 101 health plans in California MA health plans ICC: 0.72 CA health plans ICC: 0.86. ICD codes are specified. The numerator counts all emergency visits and hospitalizations with a primary or secondary ICD-based diagnosis of asthma in a child aged 3-21 years old who was eligible in the reporting month. The denominator is 100 child years for children ages 3-21 years old with identifiable asthma.
- Data elements are clear. No concerns.
- No major concerns
- More explicit definitions (e.g. denominator) would be helpful.
- Data elements are not consistently clearly defined. More clarification is needed to ensure the validity of the reporting results.
- No concerns

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

- reliability was acceptable.
- no
- testing conducted at the score level
- These results show that when assessing reliability by intraclass correlation coefficients (ICCs) at the plan level, the measure has good to very good reliability.
- No concerns.
- No major concerns
- Reliability appears to be moderate.
- No concerns with the reliability testing. Though the results indicated that there could be some issues with reliability.
- No concerns

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

- The concerns of the SMP are noted. Degree of missing data is important to note. The positioning of the Asthma diagnosis code is an important point raised by the SMP. The developer argues that pediatric pts have fewer diagnoses so asthma is less likely to appear beyond the top 1-2. Would like to see these references. The argument that exclusion of cases with asthma in the second position would exclude too many cases does not carry much weight. Would like to see evidence demonstrating that a code in the primary or second position are associated with a real diagnosis of asthma.
- It appears that there is predictive validity, in that with QI initiatives impacting asthma care, there was a reduction of ED visits. This was only in one **market,** however. There appears to be missing data in some of the

review for validity, and from my impression, how does asthma as a secondary diagnosis for ED visit factor into validity assessment?

- SMP was not able to come to consensus
- For empirical validity testing, construct validity was the primary analysis. Using California Medicaid data, the relationship between plan performance on the measure and plan performance on related and unrelated HEDIS measures was assessed. This analysis supports construct validity, demonstrating greater correlation between the proposed measure of pediatric ED use and related measures than between pediatric ED use and unrelated measures. For predictive validity testing, a difference in differences approach was used for Vermont practices in a learning collaborative focused on pediatric asthma compared to similar Vermont practices not in a learning collaborative. The analysis supports predictive validity, demonstrating that the measure is responsive to a QI initiative.
- No concerns
- Unclear validity-need further discussion
- Not sure if the weak correlation of the Compliance to the measure is what the developer **would** have expected or not. Could the developer explain the rationale for using the other measures (Low back pain imaging) as they don't make much sense for comparing the validity of avoiding the ED for asthma.
- There were inconsistencies in the validity testing across states which identified missing data elements. Risk adjustment also appeared to have variability which should be addressed.
- No concerns

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

- Relatively high number of members with missing data may limit the ability of the model to risk adjust.
- Appears appropriate
- na
- Children with a concurrent or pre-existing cystic fibrosis diagnosis or emphysema diagnosis were excluded. These exclusions led to very few patients being dropped (<2%). Children who have not been consecutively enrolled in the reporting plan within the same payer for at least three months were excluded. Initially, 12 months of continuous health plan enrollment was required for inclusion. This was to avoid patients being attributed to a health plan that had not cared for the patient for very long and thus should not be held accountable for outcomes or utilization of these patients. However, analysis showed that the requirement of continuous enrollment of at least three months is preferable, based on retention of a much larger number of patients and based on similar levels of continuous enrollment across plans with at least 50 patients. A statistical risk model with six risk factors was used a negative binomial regression model to account for the dispersed nature of the outcome. All factors were retained after using a backward selection process to eliminate variables with p>0.10, Twenty six health care plans in Massachusetts were analyzed. State-wide, there were 18.4 asthma-related ED visits/100 child-years. There were 2 health plans that moved to a higher performance ranking with social risk factor adjustment, compared to baseline risk adjusted performance. No health plans moved to a lower performance ranking with social risk factor adjustment. Health plans whose performance improved with social risk factor adjustment had patients living in zip codes with higher poverty, lower incomes, lower educational achievement, and more unemployment. These analyses demonstrate that SES factors seem to drive

performance, generally driving a change in performance for practices with lower SES patient populations. Based on these analyses, we suggest including SES risk factors in comparative performance measurement.

- Multiple threats seem to exist around secondary asthma, and missing data elements
- Unclear validity-need further discussion
- It would be helpful to hear the developer discuss the risk adjustment issues as brought up by the SMP.
- Some concerns overall about the validity of this measure.
- No concerns

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data) 2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- response above covers these components
- See response above
- na
- Standard z-score methodology was used to identify high, medium, and low performers, based on CMS scoring for quality measures. Plans with a Z-statistic >1.96 were considered poor performing outliers, those with
- Complex data analysis.
- Unclear validity-need further discussion
- Concern about social risk factors for adjusting risk.
- Missing data. Data not consistent across states/testing. Basic measure statistical testing was not provided in order to determine validity of testing.
- No concerns

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.

- Coded by someone other than person obtaining original information
- ALL data elements are in defined fields in electronic claims

Questions for the Committee:

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

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

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

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

- no concerns.

- Data collection is heavily electronic. SDoH data is harder to collect accurately in EMR or claims data
- data elements are in defined fields in electronic claim
- All data elements are in defined fields in electronic claims. Data collection seems straightforward.
- Data elements are routinely generated during care delivery. No concerns
- No major concerns
- Seems very feasible.
- Data elements should be available for this measure based on the specifications. Data should be available in electronic form.
- Data elements routinely generated in healthcare delivery

Criterion 4: Usability and Use

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

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

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

Planned use in an accountability program? ☒ Yes ☐ No

Accountability program details

- This is a new measure, and not yet in use.
- It is publicly available to all, with technical specifications posted online for public use.
- Developer plans to work with states to incorporate into Medicaid programs.

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

Feedback on the measure by those being measured or others:

- The developer shared the results of the measure during a developer convened asthma quality improvement collaborative convention for California-based pediatric primary care practices. Based on a survey conducted by the developer, most sites found the information helpful. They suggested the following improvements:
    - Include data on urgent care visits, since some clinics have an urgent care that manages asthma exacerbations most of the time.
    - The data from one of the sites was difficult to get and they were not confident in the number of patients (they seemed very low, and the site knew there were more patients for the denominator)

Questions for the Committee:

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

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

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

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

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

Improvement results:

- Measure has not been implemented. Developer did not provide year-over-year performance data.
- The developer states that improvement on this measure was associated with participation in the Vermont state-level quality improvement collaborative.

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

Unexpected findings (positive or negative) during implementation: The developer identified none.

Potential harms: The developer identified none.

Questions for the Committee*:*

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

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

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

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

- not yet being used. technical specs are posted online. Feedback was solicited during an asthma QI conference.
- Feedback from Medicaid state directors sought. Working with National Improvement Partnership Network for feedback
- not yet in use
- This is a new measure and since it is not yet endorsed, it is not currently in use. It is publicly available to all, with technical specifications and SAS code posted online for public use. Dissemination and measure uptake are key goals of the Pediatric Quality Measurement Program. AHRQ provides leadership in developing and disseminating materials to facilitate uptake across health plans and accountability programs. It is anticipated that there will be interest in implementation in at least one state, with potential use of the measure within 18-24 months. After the quality improvement collaborative was complete, team leadership conducted semi-structured qualitative interviews with physician and QI champions at each of the participating clinics. In response to the national

advisory council members, we included social determinants of health variables into the risk adjustment model, following the NQF and ASPE guidance on considerations around data sources and rationale for inclusion vs. not including these variables. We used an evidence-based approach to including these variables.

- Not yet in use. Plan to submit for inclusion in medicaid programs.
- No major concerns
- Suggestion of using for Urgent care should be explored.
- Not yet in use. Has been tested in Medicaid programs (Vermont). Results were shared with testing sites and feedback was provided to the measure developer. It is unclear if the developer acted upon the feedback.
- Planned use, not in current use

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

- improvement on this measure was associated with participation in the Vermont state- level quality improvement collaborative. No harms.
- Data will be available for review. The improvement efforts will be more labor intensive of collaborative model is a superior method to improve performance. Benefits outweigh unintended consequences.
- na
- "Improvement on this measure was associated with participation in the Vermont state-level quality improvement collaborative." But, this statement does not address the question of 12.4b1 (How can the performance results be used to further the goal of high-quality, efficient healthcare?) There were no unintended consequences.
- No harms
- No major concerns
- Appears to meet the requirements for Usability.
- Results could improve healthcare quality and efficiency. No identified harms were provided.
- No concerns

Criterion 5: Related and Competing Measures

Related or competing measures

- 0728 : Asthma Admission Rate (PDI 14)
- 1381 : Asthma Emergency Department Visits

Harmonization

- Measure #1381 is no longer endorsed.
- Measure # 0728: The developer states “Full technical specifications are not available as this measure is being reviewed for maintenance of endorsement. However, the measure we propose focuses on a different types of utilization, ED use, rather than asthma hospitalizations. Measure 0728 is also intended for population level analysis at the regional or state level, which differs from the use case for the proposed measure, which is health plan use, generally in collaboration with primary care practices.”

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

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

- There are two other measures: 1. 1381 is no longer endorsed. 2. 0728 - asthma admission rate. under review for maintenance of endorsement.
- Differing population levels will create challenges for harmonization
- two other related measures; one no longer endorsed, one different focus
- Full technical specifications for NQF #0728 : Asthma Admission Rate are not available as that measure is being reviewed for maintenance of endorsement. However, this proposed measure NQF #3599 : Pediatric Asthma Emergency Department Use focuses on ED use, rather than asthma hospitalizations. Measure 0728 is also intended for population level analysis at the regional or state level, which differs from the use case for the proposed measure, which is health plan use, generally in collaboration with primary care practices. The differences in specifications are justified.
- Related measure 0728 focusses on different outcome, hospitalization and at different level (state) rather than plan level.
- None identified
- Related measure 0728 seems to be measure something different, but it would be good to clarify this as part of the review.
- Competing measures identified were 0728: Asthma Admission Rate (PDI 14) and 1381: Asthma Emergency Department Visits. Harmonization of measures did not occur.
- No concerns

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 01/21/2021

- No NQF Members have submitted support/non-support choices as of this date.
- No Public or NQF Member comments submitted as of this date.

Combined Methods Panel Scientific Acceptability Evaluation

Evaluating Scientific Acceptability: Instructions

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3599

Measure Title: Pediatric Asthma Emergency Department Use

Type of measure: Panel Member #7: I would appreciate discussion of this.

☒ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☒ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☐ **Registry Data**

☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☐ **Facility** ☒ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☒ **New** ☐ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)
Panel Member #5: This measure was assessed previously (Spring 2020 cycle) and did not pass following some methodology issues. This submission included additional testing compared to the last submission, some that were recommended by the SMP.

RELIABILITY: SPECIFICATIONS

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

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

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

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

   Panel Member #1: No concerns.

   Panel Member #3: none

   Panel Member #5: Specifications are very clear – no concerns.

   Panel Member #6: Denominator seems unusually complex—exact method for calculation is confusing and rationale for multiplying by 1200 not clear. Numerator seems potentially misleadingly broad—any patients with ER visit or admission with primary or secondary diagnosis of asthma. Once asthma is carried on the problem list, it is likely to be listed for every ER and admission regardless of whether the visit or admission had anything to do with the asthma.

RELIABILITY: TESTING

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

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

   ☐ **Yes** ☐ **No** Panel Member #1: N/A

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

   Submission document: Testing attachment, section 2a2.2

   Panel Member #1: Developer used a split sample framework to compute an ICC which is appropriate.

   Panel Member #2: Split sample ICC – unclear which type of ICC is used here.

   Panel Member #3: Split ample analysis. ICC was 0.72 in MA and 0.86 in CA health plans. This is acceptable.

Panel Member #4: Conducted split-sample reliability testing

Panel Member #5: Developers used a split-sample reliability method using data from both states (MA + CA). Only 26 out of 29 health plans were included from MA. Can developers clarify why (or is this a typo)?

Panel Member #6: Split sample testing

Panel Member #7: Split-sample reliability testing using both MA and CA data. "We used the following approach, which was suggested to us when we first submitted this measure to NQF for endorsement: (1) the sample is randomly split into two halves, (2) the performance of each plan is estimated in each of the two data samples, and (3) the two sets of performance are then compared using the ICC." I will appreciate others' input on this.

Panel Member #8: Unfortunate for this field, there are a lot of flavors of ICC. It is confusing. But I think these developers did the right thing: "We used split-sample reliability testing using both MA and CA data. We used the following approach, which was suggested to us when we first submitted this measure to NQF for endorsement: (1) the sample is randomly split into two halves, (2) the performance of each plan is estimated in each of the two data samples, and (3) the two sets of performance are then compared using the ICC." My only question is which ICC they are reporting. Although mapping this context on the existing literature is tricky, they should be using an ICC that comes from a one-way model and agreement (vs consistency). If single rater ICC was used, the results can be corrected to estimate the reliability of the whole sample.

Koo, Terry, and Mae Li. 2016. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research." *Journal of Chiropractic Medicine* 15 (March). doi:10.1016/j.jcm.2016.02.012.

https://www.datanovia.com/en/lessons/intraclass-correlation-coefficient-in-r/

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

   Submission document: Testing attachment, section 2a2.3

   Panel Member #1: The results indicate reasonable level of reliability (ICC>0.70) in both samples of health plans.

   Panel Member #4: Results show that when assessing reliability at the plan level, that the measure has good to very good reliability (ICC range from 0.72 – 0.86)

   Panel Member #5: It would be useful to demonstrate the range or reliability for all included health plans along with their sample size, or at least percentile-based ICC values. I suggest this be added to the submission as additional material before the SMP evaluation meeting, if possible.

   Panel Member #6: A little unclear but seems like the data from each health plan was randomly split and then the test results for each half of each plan were measured for ICC with high rates of agreement: 0.72 for MA and 0.86 for CA. The question that I have is this: calculation is complex and involves many steps. Measure steward is not clear and therefore it is unknown if healthplans are expected to report this information or whether some other entity, such as CMS, will have access to all of the data necessary and generate the reports. If the latter, the testing performed would indicate high potential reliability. If the former, then the testing is inadequate.

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

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

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

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

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

☐ **No**

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

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

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

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

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

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

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

    Panel Member #1: See comments for item #7. More information about the reliability of plans (median estimate, % of plans with ICC > 0.70)

    Panel Member #2: The reliability values are good, but the unclear if the ICC form used was appropriate here. Could be deemed as 'high' if that was documented.

    Panel Member #3: ICC was 0.72 in MA and 0.86 in CA health plans.

    This is acceptable.

    Panel Member #4: Used appropriate testing method; found good to very good reliability in the two data sets.

    Panel Member #5: The moderate rating is due the moderate levels of reliability for the MA health plans.

    Panel Member #6: Greater clarity regarding the definition of the denominator as well as greater clarity regarding who/what entity is going to be generating the report.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

    Submission document: Testing attachment, section 2b2.

    Panel Member #1: None.

    Panel Member #3: none.

    Panel Member #5: Exclusions reported were tested using only data from MA.

    This issue has been raised during the last review of this measure, and developers responded with additional information on exclusion rates for CA (which were very small). Please include this additional information in this submission!

    Panel Member #6: Concern related to numerator in section 2 above; is what is being measured truly related to asthma management?

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

    Submission document: Testing attachment, section 2b4.

    Panel Member #1: None.

    Panel Member #2: None

    Panel Member #3: none

    Panel Member #4: None. See meaningful variation across plans, with 40%-45% of plans identified as high or low performing

Panel Member #5: No concerns. Rates of Asthma ED visits could be added to the CA table.

Panel Member #6: Concern related to numerator in section 2 above; is what is being measured truly related to asthma management?

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Submission document: Testing attachment, section 2b5.
Panel Member #1: N/A
Panel Member #2: None
Panel Member #3: none
Panel Member #4: Not applicable.
Panel Member #6: Performance of the model was dramatically different in the derivation and testing data sets

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

Submission document: Testing attachment, section 2b6.

Panel Member #1: None.

Panel Member #2: None

Panel Member #3: none

Panel Member #4: It is still unclear the **impact** of the missing SES data for plans that have substantial missing data (10%+).

Panel Member #5: No concerns.

Panel Member #6: Certain elements have high level of missingness, which may partially account for the difference in model performance between Medicaid and all payer data **Risk Adjustment**

16a. **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

Panel Member #3: Negative binomial regression model.

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☒ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☒ Yes ☐ No ☐ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☒ Yes ☒ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☐ No

Panel Member #6: Not directly discussed

16d. **Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☐ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☐ Yes ☐ No

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No

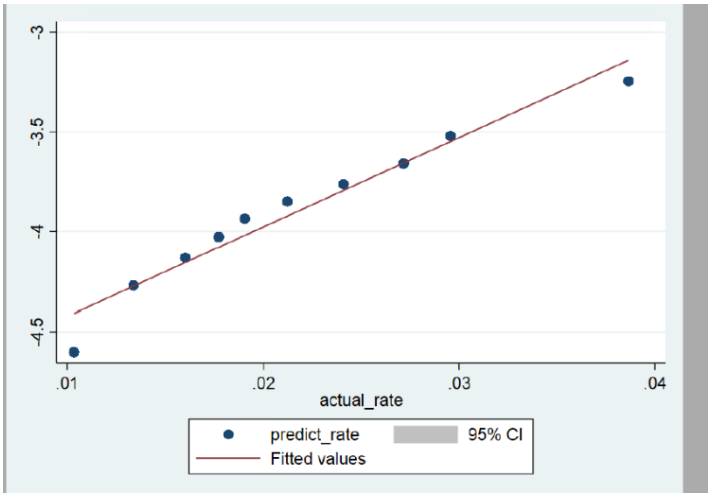
16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☒ No

Panel Member #6: R2 highly variable between development and validation sets. Confidence intervals and p values not given. Very few variables (6) included in the model. Method of variable selection was a priori and not well explained.

16d.5. Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☐ No

16e**. Assess the risk-adjustment approach** Panel Member #1: The c-statistics reported indicates very good model discrimination. The calibration statistics support the validity of the model.

Panel Member #3: Model calibration is very good in the validation data set.



Panel Member #4: Concerns with the low r-squared values for the CA plans (13%), which is very different value from the r-squared value for the MA plans (56%). Could it be a case that the model is a better fit for APCD than Medicaid only? The Medicaid only population may have less variation in the community risk factors (% below poverty, education, unemployment) which may make the models less predictive.

Panel Member #5: No concerns

Panel Member #7: "The c-statistic is not appropriate for this model since the measure is a rate, not a proportion.

We assessed risk calibration by calculating the expected and predicted mean rates across deciles of patients, calculating a correlation coefficient using the pwcorr command in STATA (pairwise correlation coefficient), and creating visual plot to examine the plot of predicted vs. actual across deciles."

R-squared values:

MA data: 0.56

CA data: 0.13

Although calibration plots look good, I have concerns about the adequacy of the risk adjustment, particularly the variation by state.

***For cost/resource use measures ONLY*:**

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

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

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

VALIDITY: TESTING

18. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**
19. **Method of establishing validity of the measure score:**

☐ **Face validity**

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

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

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

Submission document: Testing attachment, section 2b2.2

Panel Member #1: The developer's strongest tests are correlating measure performance with plan performance on related and unrelated measures and they further used a difference-in-difference model to estimate QI impacts on measure score.

Panel Member #2: Construct validity assessed by comparing to HEDIS metrics.

Panel Member #3: Construct validity: assessed correlation between performance on measure and performance on related and unrelated HEDIS measures

- Low correlation with HEDIS measures

- Assessed the association between statewide QI effort and decrease in ED utilization using DD analysis

- QI initiative was associated with decrease in ED utilization

Panel Member #4: Used CA Medicaid data and assessed the relationship between plan performance on the measure and plan performance on related and unrelated HEDIS measures. Related HEDIS measures included: Medication management for asthma; and vaccines in children. Unrelated measures included ACE monitoring, Cervical CA screening, and imaging for low back pain.

Panel Member #5: Empirical validity at the health plan level was presented, by assessing correlations between the proposed measure of Pediatric ED use with a set of related and unrelated measures.

Panel Member #6: Correlation with various other measures

Panel Member #7: Construct validity

"We used CA Medicaid data and assessed the relationship between plan performance on the measure and plan performance on related and unrelated HEDIS measures. Related HEDIS measures included: Medication management for asthma; and vaccines in children. Medication management measures for asthma are the most closely related to asthma ED use, but are not reported separately for pediatric patients in HEDIS, so we would not expect perfect correlation. Therefore, we also looked at HEDIS pediatric vaccine status, which we would expect to be higher in health plans focusing on pediatric quality, including potentially pediatric asthma. Our unrelated measures included ACE monitoring, Cervical CA screening, and imaging for low back pain."

Predictive Validity

"difference in differences approach for practices in a learning collaborative focused on pediatric asthma compared to similar practices not in a learning collaborative in VT"

Panel Member #8: **"**Calculation of the relationship between plan performance on the measure and plan performance on related and unrelated HEDIS measures. Related HEDIS measures included: Medication management for asthma; and vaccines in children. Medication management measures for asthma are the most closely related to asthma ED use, but are not reported separately for pediatric patients in HEDIS, so we would not expect perfect correlation." I like that they correlated the outcome measure with process measures that should impact the measure score. If patient-level data were available, that would be even stronger (i.e.., patients receiving guideline concordant care have fewer ER visits) if analyzed in a multilevel model. I like the logic behind the DnD analysis… basically checking if in-hospital QI efforts can change the measure score.

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

Submission document: Testing attachment, section 2b2.3

Panel Member #1: The testing results showed that the ED admissions are sensitive to QI initiatives and that the pedi-ed measure is correlated with related quality measures.

Panel Member #2: Correlation is very low with 2 of the 3 related measures.

Panel Member #4: Found stronger correlations between the proposed measure of Pediatric ED Use and related measures than between Pediatric ED use and unrelated measures

Panel Member #5: Results supported the hypothesized relationships, demonstrating greater correlation between the proposed measure and related measures (0.1-0.3) than between Pediatric ED use and unrelated measures (r<0.1). Although correlations with related measures were not strong, developers explain why they would not expect them to be strong, which makes sense. Only one correlation with a related measure was significant (correlation with Child Vaccines, r=0.33, p=0.02).

Panel Member #6: Correlation strongest with HEDIS measure for asthma medication (process measure)

Panel Member #7: This is a challenging measure for validity assessment. Given this challenge, the approach is fine.

Panel Member #8: I agree with the developers interpretation of the validity analysis. The results provide some rough signal of validity. I especially like the effort to match process measures with outcome measures. Note, these results could suffer from the ecological fallacy and patient-level analysis would me much stronger.

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

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☒ **No**

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

    Panel Member #6: Given that secondary diagnosis for asthma may be unrelated to the reason for ER visit/admission, the validity of the measure as constituted has not been established.

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

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

    ☐ **Yes**

    ☐ **No**

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

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

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

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

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

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

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

    Panel Member #1: See comments in #22

    Panel Member #2: Low construct validity values.

    Panel Member #3: Risk adjustment model demonstrates acceptable model calibration.

    Results of empiric validity analysis are acceptable.

    Panel Member #4: Concerns remain with the number of plans that have a relatively high percentage of members who are missing social risk factor data. The social risk factors are key adjustment variables in the risk adjustment models. 20,000 of 85,000 members are in plans for which 10% of more of members are missing social risk data.

And concerns remain with the risk-adjustment model. Given the possible differences in SES factors for APCD and Medicaid populations, applying a singular risk model to all populations may prove challenging/problematic.

Panel Member #5: The moderate rating is due to moderate levels of validity demonstrated.

Panel Member #6:

1) Given that secondary diagnosis for asthma may be unrelated to the reason for ER visit/admission, the validity of the measure as constituted has not been established.

2) Disparity in performance of predictive modeling between CA and MA is concerning

Panel Member #7: I could score moderate. Variation in risk adjustment is a concern.

Panel Member #8: See above responses

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

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

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

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

ADDITIONAL RECOMMENDATIONS

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

Developer Submission

Additional evaluations and submission materials attachments…

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

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

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

3599_NQF_evidence_attachment_2020_11_20.docx

1a.1 *For Maintenance of Endorsement:* Is there new evidence about the measure since the last update/submission? Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 3599

Measure Title: **Pediatric Asthma Emergency Department Use**

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

Date of Submission: **11/8/2020**

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

Outcome

☒ Outcome: **Pediatric asthma-related emergency department utilization**

☐ Patient-reported outcome (PRO):

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

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

☐ Process:

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

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

Please see next page for the **Key Driver Diagram (KDD)** informing the relationship between processes and the measure outcome—pediatric asthma-related ED visits. This KDD was developed to support the state-wide IMPLEMENT Asthma learning collaborative in California, under the PQMP program.

In addition, the logic model underlying the measure development includes additional drivers related to the social determinants of health (SDOH). These include components related to the social determinants noted in the NQF guidance on whether to risk adjust for these variables in outcome measures: education, income, employment.

Additional specific components to the proposed measure:

--Family and patient health literacy, enabling understanding and ability to follow asthma management recommendations, and decrease pediatric asthma utilization. [Education]

--Ability to access and afford asthma medications, leading to decreased pediatric asthma utilization through better asthma control and prevention of exacerbations. [Income, Employment]

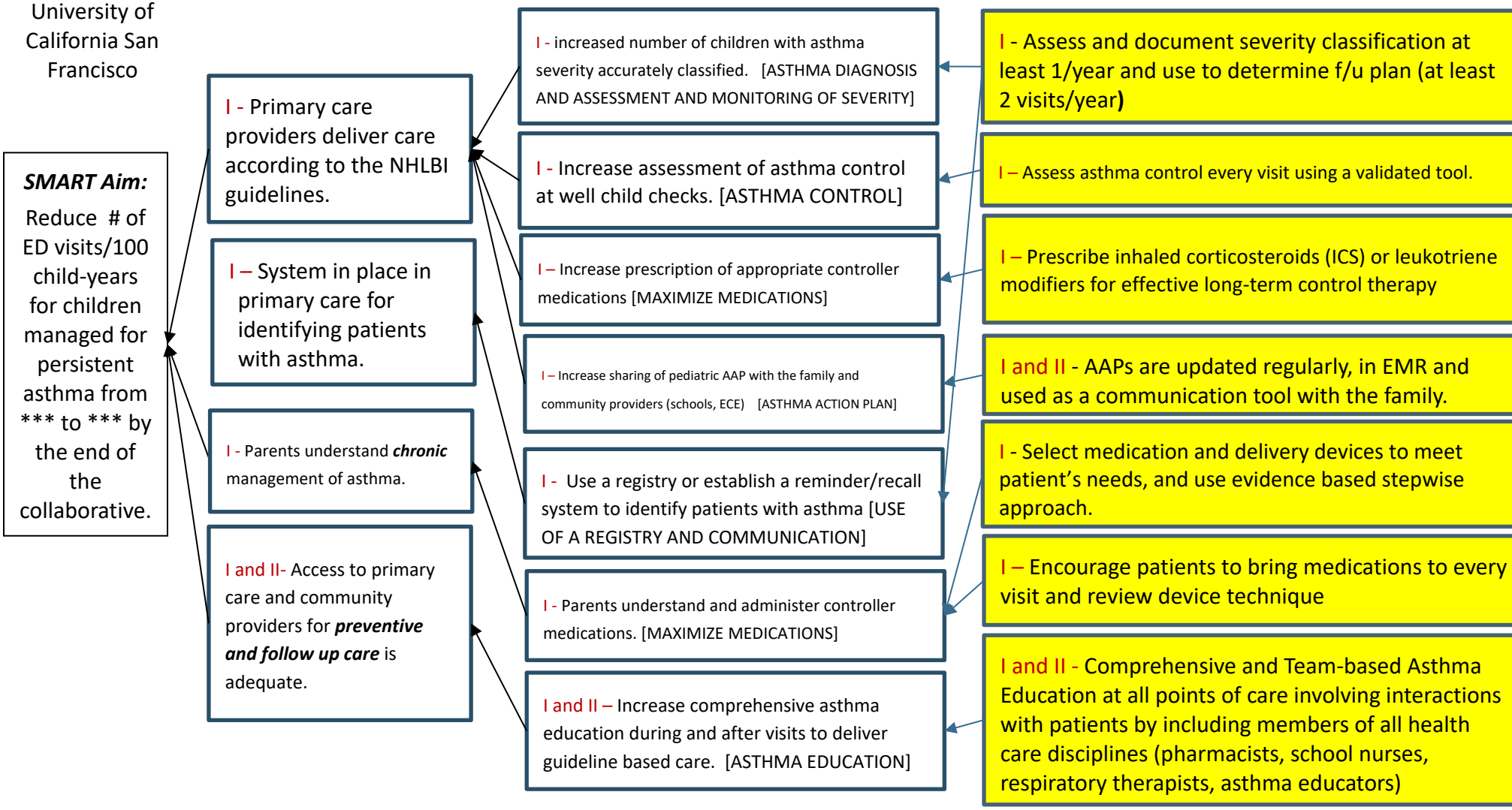
--Safe and stable housing with ability to limit asthma allergen triggers including mold, dust, second hand smoke, leading to decreased pediatric asthma utilization due to decreased acute asthma exacerbations. [Income, Employment]

The variables used from the ACS for the risk-adjustment model have been validated as measures of SES using factor analysis and found to be associated with increased readmissions as well as direct measures of allostatic load or physiological stress.[1,2] This makes them particularly well suited for inclusion in this measure of utilization.

Of note, since race/ethnicity are not direct measures of social risk, it is not recommended per NQF or ASPE guidance to use them for social risk factor adjustment.

1. Martsolf GR, Barrett ML, Weiss AJ, et al. Impact of Race/Ethnicity and Socioeconomic Status on Risk-Adjusted Readmission Rates: Implications for the Hospital Readmissions Reduction Program. *Inquiry : A Journal of Medical Care Organization, Provision and Financing.* 2016;53:0046958016667596.
2. Bird CE, Seeman T, Escarce JJ, et al. Neighbourhood socioeconomic status and biological 'wear and tear' in a nationally representative sample of US adults. *J Epidemiol Community Health.* 2010;64(10):860-865.

University of California San Francisco



Authors: Keith Robinson, MD ; Christine Pellegrino, MS; Judy Shaw, EdD, MPH, RN FAAP, University of Vermont

Updated 1/24/18

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

NA

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

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

This measure was refined and further tested in the Pediatric Quality Measures Program, funded by the Centers for Medicaid and Medicare Services and administered through the AHRQ.

As part of this work we assessed the relationship between improved performance on specific asthma care processes, achieved through a state-wide quality improvement collaborative in Vermont, and decreased asthma ED visits, using the specifications for the proposed measure #3599 of pediatric asthma ED visits.

The paper is in press at *Pediatrics*, under embargo until November 23, 2020. In consultation with the journal, we present the abstract below and the link to the electronic version of the paper here: www.pediatrics.org/cgi/doi/10.1542/peds.2020-0213. The electronic link will be active starting Nov 23 at 12:01am.

The key table to review for additional presentation of relevant numerical results is Table 4 and is provided below.

This study provides strong evidence that improvement in asthma-related process measures is associated with improvement in performance on the proposed measure of pediatric asthma-related ED visits.

Title: Statewide Asthma Learning Collaborative Participation and Asthma-related Emergency Department Use

Background

Quality improvement (QI) efforts can improve guideline-recommended asthma care processes in the pediatric office setting. We sought to assess whether practice participation in an asthma QI collaborative was associated with decreased asthma-related emergency department (ED) visits.

Methods

A statewide network of practices participated in a pediatric asthma QI collaborative from 2015-2016. We evaluated asthma-related ED visit rates/100 child-years for children ages 3-21 years with asthma using the state's all-payer claims database. We used a difference-in-differences approach with mixed-effects negative binomial regression models controlling for practice and patient covariates. Our main analysis measured the outcome before (2014) and after (2017) the QI collaborative at fully participating versus control practices. Additional analyses assessed a) associations during the intervention period (2016) and b) associations including practices partially participating in QI collaborative activities.

Results

In the post-intervention year (2017), participating practices' (n=20) asthma-related ED visit rate decreased by 5.8/100 child-years compared to an increase of 1.8/100 child-years for control practices (n=15; difference-in-differences=-7.3, *P*=.002). Within the intervention year (2016), asthma-related ED visit rates decreased in participating practices compared to controls but were not statistically significant (-4.3 difference-in-differences, *P*=.17). The analysis including partially participating practices yielded similar results and inferences to our main analysis.

Conclusions

Participation in an asthma-focused QI collaborative was associated with decreased asthma-related ED visit rates. For those considering implementing this type of QI collaborative, our findings indicate that it takes time to see measurable improvements in ED visit rates. Further study is warranted regarding QI elements contributing to success for partial participants.

Table: Comparison of participating and control practice mean asthma-related emergency department visit rates from baseline and post-collaborative years (main difference-in-differences analysis)

Main Analysis	Asthma-related ED visits per 100 child-years	*P*-value	95% Confidence Interval: Lower	95% Confidence Interval: Upper
Control 2014	16.97	*	11.99	21.95
Participating 2014	15.29	*	11.62	18.96
Control 2017	18.80	*	12.89	24.71
Participating 2017	9.50	*	7.07	11.93
Difference-in-Differences	*-7.62*	0.002	-13.45	-1.78

Notes: Marginal rates are adjusted for patient (age, sex, and insurance) and practice (specialty, organizational structure, size, geographic region, and federally qualified/rural health center) characteristics and also for patients clustering within practices; SE: standard errors are estimated using the delta method; Coefficients for relative rates are in Supplemental Table 2.

*cell intentionally left blank

Processes measured:

The specific asthma care processes of focus in the VT collaborative were based on NHLBI guidelines for asthma management. They included: 1) The percentage of patients with asthma severity documented as intermittent or persistent (mild, moderate, or severe); 2) The percentage of patients prescribed inhaled corticosteroids or other control medications if asthma severity is persistent; 3) The percentage of patients with asthma control assessed with a validated tool; 4) The percentage of patients with an asthma action plan initiated, reviewed or updated as needed within the last 12 months; 5) The percentage of patients with at least one planned asthma visit every 6 months; 6) The percentage of patients assessed for tobacco use/exposure; 7) The percentage of patients and their caregivers educated about their asthma; 8) The percentage of patients and their caregivers instructed on how to use their asthma delivery device; and 9) The percentage of patients between the ages of 5–21 who completed a spirometry test in the past 24 months. Measures 1–4 were required and teams chose any two of the remaining five to improve at their practice.

Publication of Processes and Collaborative Implementation: Additional description of the collaborative and the process measures used were published in the Journal of Asthma in Dec 2019. Citation for this additional paper: Weinberger et al. A primary care learning collaborative to improve office systems and clinical management of pediatric asthma. *J Asthma*. Dec 2019 14;1-10.

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

What is the source of the *systematic review of the body of evidence* that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but

separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

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

☐ Other

Systematic Review	Evidence
Source of Systematic Review: • Title • Author • Date • Citation, including page number • URL	*
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?	*

*cell intentionally left blank

1a.4 OTHER SOURCE OF EVIDENCE

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

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

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

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

1b. Performance Gap

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

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

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

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

In 2009, Congress passed the Children´s Health Insurance Program Reauthorization Act (CHIPRA, Public Law 111-3), which presented an unprecedented opportunity to measure and improve health care quality and outcomes for the nation´s children, including those enrolled in Medicaid/CHIP. When CHIPRA was enacted, the Agency for Healthcare Research and Quality (AHRQ) and the Centers for Medicare & Medicaid Services (CMS) began working together to implement selected provisions of the legislation related to children´s health care quality.

The law called for the establishment of the CHIPRA Pediatric Quality Measures Program (PQMP) to improve and strengthen the "Child Core Set" of measures and develop new measures as needed.

The proposed measure 3599 Pediatric Asthma Emergency Department Use was developed and then further tested and refined under the PQMP. Here we present the rationale for the proposed measure, informed by our work in convening state-level quality improvement collaboratives (one in CA and one in VT) specifically focused on improved asthma care, and assessing the relationship between improvements in asthma care processes and performance on the proposed pediatric ED utilization measure. The following text draws from the PQMP Toolkit for the measure for use by clinics and health plans interested in using the measure for quality improvement efforts.

_____

What is needed for QI success:

The Pediatric Asthma Emergency Department Use measure is an outcome measure based on administrative data. In that context, the toolkit user entity (state agency, health plan, healthcare organization, improvement partnership, provider group) will need to partner with practices and quality improvement coaches to evaluate systems and develop process measures to guide improvement efforts that will impact the measure. Examples of process measures can be developed from clinical guidelines such as the National Heart Lung and Blood Institute Guidelines for the Diagnosis and Management of Asthma.

It is important to recognize that most process measures will be contained in a practice's electronic health record (EHR) or a data source separate from claims data. Additionally, effective process measures should be evaluated longitudinally to assess performance over time and allow for identification of variation, either intended or unintended.

Successful improvement requires sound quality improvement science methodology, appropriate resources and ready access to reliable data. Without these components (appropriate training, infrastructure and data access), application of QI may lead to unintended consequences, such as provider frustration or QI 'fatigue'.

See below for a summary of potential strategies to support implementing quality measurement and improvement strategies in primary care settings to reduce asthma-related ED visits from the perspective of a health plan.

Summary of Strategies and Complementary Toolkit Resources

1) Goal: Understand the population and the system resources in your care delivery area
    a. Resources Required: Understand the population and the system resources in your care delivery area
    b. Health Plan contributions: Foster partnerships and determine strategic alignment(s)
2) Goal: Partner with practices in Health Plan network
    a. Resources Required: Practice network
    b. Health Plan contributions: Engage practices in collaborative
3) Goal: Engage practice leadership
    a. Resources Required: Practice champion
    b. Health Plan contributions: Financial alignment for clinical champion(s)
4) Goal: Develop improvement science expertise
    a. Resources Required: QI Coaching
    b. Health Plan contributions: Offer financial support for QI infrastructure
5) Goal: Determine baseline performance on NHLBI measures
    a. Resources Required: Process measures from EHR
    b. Health Plan contributions: Support practices to engage EHR vendor/ practice support to obtain data
6) Goal: Assess periodic performance/improvement over time
    a. Resources Required: Periodic data pull from EHR for process measures (by practice-based clinicians or chart auditor)
    b. Health Plan contributions: Develop practice-based incentives for improvement
7) Goal: Understand variation in performance and guide improvement efforts
    a. Resources Required: Practice level strategies
    b. Health Plan contributions: NA
8) Goal: Systems Learning
    a. Resources Required: Practice data of children who went to ED
    b. Health Plan contributions: Health plan provides practice reports on ED utilization for clinic health plan members

QI Strategies

Overview: Approaches to Quality Improvement in Asthma Care

There are numerous factors and settings that impact the asthma emergency department (ED) measure (e.g., schools, ED, acute care, access to specialists, community, etc.), and must be considered in trying to reduce inappropriate ED use for pediatric asthma. Many factors can lead to a child with asthma receiving care in the ED such as poor asthma control, severity of symptoms, decreased access to care, and ability to enact emergency care (such as use of a rescue inhaler) among many others. When thinking of these factors and

where they occur, they generally can be attributed to the patient's home and school environment, medical home, the ED or a combination (Allen, 2019). Interventions engaging the ED should be considered if there is a high rate of patients with multiple visits to the ED. In this scenario, it is important to evaluate access to care, environmental factors, ED care and the connection between the ED and the medical home.

There are three general quality improvement (QI) approaches to decrease pediatric ED visits for asthma that have a strong evidence base:

- Primary Care,
- Provider Continuing Medical Education, and
- Parental and School-Based

There is insufficient evidence to recommend a single approach, or set of interventions, over another because there are many factors that influence what will be the most effective approach for a care system. Some systems with a high degree of integration and QI capacity have chosen multiple interventions to reduce ED visits (Allen, 2019). However, most care systems will likely be best served to identify a single approach after evaluating their outcome and process measures while identifying the key drivers of performance. Assembling an interprofessional team to understand key stakeholder priorities and readiness coupled with a thorough and systematic approach to QI are essential to achieving success.

This pediatric asthma measure has potential to improve asthma care, reduce ED utilization, and promote collaboration between health plans and primary care practices. Successful utilization of the measure will necessitate interpreting data from multiple sources and business entities. Because of this, there will be practical, ethical and legal limitations relative to sharing data and how improvement efforts are implemented. While the approaches described above each have merit, the PQMP grantees charged with testing how to use the pediatric ED use measure chose to focus on the intervention area with the most evidence of success. This toolkit outlines primary care-focused interventions using an intensive educational approach and methods to develop improved systems of care.

Primary Care-Focused Approach

Most interventions that have been successful in improving asthma ED outcomes through provider-based activities have included intensive educational approaches or methods to develop improved systems of care within the primary care office setting. Harder et al. examined the effects of a one-year QI collaborative for primary care clinicians that focused on office systems strategies (e.g. asthma assessment, control and management, and patient education). Compared to control practices, the participating practices noted a substantial decrease of nearly 40 percent in asthma-related ED visit rates more than a year after the end of the collaborative (Harder, 2020). The development of a systematic primary care approach to asthma care can also improve asthma health care utilization. In a pragmatic, cluster randomized controlled trial, Yawn et al. demonstrated that the use of Asthma APGAR (Activities, Persistent, triGGers, Asthma medications, Response to therapy) tools improved rates of asthma control and reduced asthma-related ED and urgent care visits (Yawn, 2018).

PQMP Toolkit Approach: Primary Care Collaboratives

The IMPLEMENT for Child Health initiative (IMPLEMENT) is the overall program that tested out the usability of the PQMP asthma ED measure by conducting QI initiatives in both San Francisco, California (SF Collaborative) and in Burlington, Vermont (VT Collaborative), both aimed to improve pediatric asthma care delivered in a primary care setting. The strategies described in this toolkit reflect the learnings from those two QI initiatives aimed at examining the usability of the asthma measure. In the SF Collaborative, primary care practices participated in a 12-month learning collaborative. In the VT Collaborative, practices had participated in an earlier Vermont statewide asthma learning collaborative (CHAMP Learning Collaborative, for more information see https://www.med.uvm.edu/vchip/champ) and therefore a more targeted approach was undertaken – performing a "deep dive" to examine factors that contributed to high ED rates. Staff and faculty from the

University of Vermont's Vermont Child Health Improvement Program's (VCHIP) provided the QI expertise for both initiatives.

_____

In summary, the proposed pediatric asthma measure is responsive to improvements in QI process measures, as demonstrated by Harder et al. (publication in press), and improvements can be driven either by individual clinics, clinics participating in a collaborative, or health plans supporting clinics in improvement efforts. Health plans could also consider addressing the social determinants of health, as described in the Logic Model section of the Evidence attachment. Addressing the social determinants of health, while not a focus of the PQMP work, is another avenue for potential intervention to improve performance on this measure.

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

CA health plans

Year: Measurement year 2016, data used from 2015-2016

Number of plans: 103

Number of patients: 321,072

Mean: 24.4

Std dev: 9.4

Min: 7.6

Max: 63.5

IQR: 18.3-28.9

Scores by decile:

CALIFORNIA

Decile	Predicted #ED visits/100 child-years
1	12.1
2	16.8
3	19.3
4	21.4
5	23.5
6	25.6
7	27.9
8	30.9
9	35.5
10	46.7

MA health plans

Year: Measurement year 2015, data used from 2014-2015

Number of plans: 29

Number of patients: 83,577

Mean: 12.7

Std dev: 6.7

Min: 0

Max: 27.7

Median: 11.2

IQR: 9.6-18.0

Scores by decile:

MASSACHUSETTS

Decile	Predicted #ED visits/100 child-years
1	5.7
2	7.2
3	10.3
4	13.5
5	17.4
6	19.8
7	24.0
8	26.0
9	29.8
10	36.4

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 (4b1) under Usability and Use.*

For CALIFORNIA DATA:

Gender:

Females: 26.0 per 100 child-years

Males: 26.1 per 100 child-years

Race/Ethnicity	Pediatric ED visits/100 child-years
White	23.6
Latinx	24.2
Black	40.6
API	15.2
Other	24.3
Unknown	35.9

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

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

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

https://chipper.ucsf.edu/studies/implement/documents

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**:** IMPLEMENT_Asthma_ED_Use_ICD_and_CPT_Codes-637413960397551146.xlsx

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

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

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

Not an instrument-based measure

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

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).*

Number of asthma-related ED visits

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 details: The numerator counts all emergency visits and hospitalizations with a primary or secondary ICD-based diagnosis of asthma in a child who was eligible in the reporting month. The asthma ICD codes are in the Excel workbook in S.2b. 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 either an ED visit or a hospitalization. In the datafiles created for the measure, the data is in member-month rows. Thus the numerator is the number of visits for that member in each month. See S.14 for more information on measure calculation.

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

100 Child Years for children with identifiable asthma

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 represents the person-time experience among eligible children with identifiable asthma (definition below). Assessment of eligibility is determined for each child monthly. The total number of child months in the measurement year experienced is summed and divided by 1200 to achieve the units of 100 child years for the denominator.

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, including the month being assessed.

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: Cystic Fibrosis (CF) diagnosis, or Emphysema diagnosis.

Please see attached list of ICD codes ("IMPLEMENT Asthma ED Use ICD and CPT Codes") for exclusion criteria for CF and emphysema.

Consecutive enrollment is defined as being consecutively enrolled within the same payer. This allows for a change in plan type (e.g. changing to a PPO to an HMO within same payer). Continuous enrollment does not include moving payers even if continuously enrolled (e.g. moving from Kaiser to Blue Cross within the three month window would exclude them from the denominator. This is due to the measure being a health plan-level measure.

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 is not a stratified measure.

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.)*

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 3 – 21 years who meet the criteria for Identifiable asthma - and do not satisfy one of the exclusion criteria - 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 2016, first one would identify children with Identifiable asthma using the criteria, and analyze all of calendar year 2015 when doing so. Continuous enrollment criterion requires that the child was enrolled in November and December of 2015, as well as January 2016. 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 2015 AND January 2016 when doing so. Continuous enrollment criterion requires that the child was enrolled in December 2015 and January 2016, as well as February 2016. 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 2015 AND January through November 2016 when doing so. Continuous enrollment criterion requires that the child was enrolled in October 2016 and November 2016, as well as December 2016. 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:

c. Prior hospitalization with asthma as primary or secondary diagnosis

d. Other qualifying events after the fifth birthday (age is age at occurrence):

ii. One or more prior ambulatory visits with asthma as the primary diagnosis, OR

iii. Two or more ambulatory visits with asthma as a diagnosis, OR

iv. One ambulatory visit with asthma as a diagnosis AND at least one asthma-related prescription

c. Other qualifying events, any age:

i. Three or more ambulatory visits with diagnosis of asthma, OR

ii. Two or more ambulatory visits with a diagnosis of asthma AND one or more asthma- related prescriptions

Note, these age differences are per NHLBI guidelines (https://www.nhlbi.nih.gov/health-topics/guidelines-for-diagnosis-management-of-asthma) and were reviewed and developed in collaboration with the Delphi panel of experts convened during the development of this measure.

Step 3. Calculate rate as Numerator / Denominator.

- If a qualified member has no numerator events during a month, the event count value is 0.

See document at https://chipper.ucsf.edu/upload/chipper/documents/Flowsheet_Asthma_1.pdf

for a flow chart for data flow and management steps to calculate the measure.

SAS code is available at https://chipper.ucsf.edu/upload/chipper/documents/asthma_1_sas_code.pdf

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

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

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.)*

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

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

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

Administrative claims, including state Medicaid claims and state All-payer claims databases.

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)*

Outpatient Services

If other:

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

NA

2. Validity – See attached Measure Testing Submission Form

Asthma_1_NQF_testing_attachment_2020_11_19.docx

2.1 *For maintenance of endorsement*

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

2.2 *For maintenance of endorsement*

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

2.3 *For maintenance of endorsement*

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

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

Measure Number (*if previously endorsed*)**:**
Measure Title: Pediatric Asthma Emergency Department Use
Date of Submission: **8/3/2020**

Type of Measure:

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

*cell intentionally left blank

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

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

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

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims

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

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

California (CA) Medicaid Claims, Massachusetts (MA) all-payer claims data.

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

CA (2016 measurement year, and 2015 look back year) Medicaid Claims, MA all-payer claims data (2015 measurement year and 2014 look back).

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

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

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

We analyzed data from CA and MA for health plans, as this is the proposed level of measure implementation.

In CA Medicaid data there were claims from a total of 103 managed health plans and the fee-for-service Medicaid program.

In MA all-payer claims data there were a total of 29 health plans.

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*)

Ages: Children aged 3 -21 years. Enrollment criterion: Children who have been enrolled for three consecutive months including the month being assessed. Event/Diagnosis: Children whose claims meet the eligibility criteria for identifiable asthma, as described in submission.

Description of Eligible Children in CA 2016 Medicaid claims

Measures	Eligible Patients, N (%)
Total	321,072 (100%)
Age group, years	*
3-5	46,095 (14.4%)
6-11	145,078 (45.2)
12-17	93,989 (29.3)
18-21	35,910 (11.2)
Gender	*
Male	181,626 (56.6%)
Female	139,446 (43.4)
Race/Ethnicity	*
White	47,126 (14.7%)
Hispanic	191,482 (59.6)
Black	37,824 (11.8)
Asian/Pacific Islander	19,009 (5.9)
Other	12,155 (3.8)
Missing or Declined to state	13,476 (4.2)
Insurance type	*
Medicaid managed care	304,358 (94.8%)
Medicaid FFSc	16,714 (5.2)
PMCAd	*
None (non-chronic)	204,333 (63.6%)
Chronic, non-complex	69,089 (21.5)
Complex chronic	47,650 (14.8)
Health Plans	*
Number of Health Plans, N	103
Number of eligible patients per health plan, Mean (SD), Median (IQR)	Mean: 3085.57 (6577.90) Median: 1094 (25%ile: 188, 75%ile: 3194.5)

[a]PPO: Preferred provider organization; [b]HMO: Health maintenance organization; [c]FFS: Fee-for-service; [d]PMCA-Pediatric Medical Complexity Algorithm, excluding from the algorithm diagnoses of asthma.

*cell intentionally left blank

Description of Eligible Children in MA 2015 APCD claims

Measures	Eligible Patients, N (%)
Total	83,577 (100%)
Age group, years	*
3-5	13,103 (15.7%)
6-11	32,717 (39.1)
12-17	24,897 (29.8)
18-21	12,860 (15.4)
Gender	*
Male	45,100 (54.0%)
Female	38,477 (46.0)
Insurance type	*
Commercial PPOa	4,348 (5.2%)
Commercial HMOb	12,004 (14.4)
Medicaid managed care	19,741 (23.6)
Medicaid FFSc	12,034 (14.4)
Other insurance type	33,806 (40.4)
More than one insurance type	1,644 (2.0)
PMCAd	*
None or non-chronic	40,232 (48.1%)
Chronic, non-complex	24,860 (29.7)
Complex chronic	18,485 (22.1)
Health Plans	*
Number of Health Plans, N	29
Number of eligible patients per health plan, Mean (SD), Median (IQR)	Mean: 2957 (4972.32) Median: 228 (25%ile: 40, 75%ile: 4180)

[a]PPO: Preferred provider organization; [b]HMO: Health maintenance organization; [c]FFS: Fee-for-service; [d]PMCA-Pediatric Medical Complexity Algorithm, excluding from the algorithm diagnoses of asthma.

*cell intentionally left blank

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: We used both CA and MA data to assess reliability.

Validity: We tested validity at the health plan level in two ways, using CA Medicaid and MA APCD data. In addition, in a secondary analysis we tested validity at the clinic level using VT APCD data. See below for descriptions of validity testing.

Risk adjustment: We developed the risk adjustment model in MA APCD data and tested it in CA Medicaid data.

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

We used social risk factors at the 5-digit zip code level, from the American Community Survey 2015 data. See listing and rationale for choice below in Risk Adjustment section.

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*)

We used intraclass correlation coefficients (ICCs) to assess 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 ICC assesses the ratio of between-site variation to within-site variation on performance. Higher ICC implies that the between site variation (signal) is higher than the within site variation (noise).[1-3] We used split-sample reliability testing using both MA and CA data. We used the following approach, which was suggested to us when we first submitted this measure to NQF for endorsement: (1) the sample is randomly split into two halves, (2) the performance of each plan is estimated in each of the two data samples, and (3) the two sets of performance are then compared using the ICC. For step 2, performance estimation was calculated using the risk adjustment approach described below, in a mixed effect model. Please see section S.14 for additional information on how to calculate the measure.

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 1. Reliability ICC testing using split sample analysis and ICC calculation of plan performance for split samples.

Level of testing	ICC	Confidence interval	Number of clusters	Number of patients	Number of patient-months*
MA Health Plans	0.72	0.49-0.86	26 plans	83,577	698,420
CA Health plans	0.86	0.79-0.90	101 plans	321,072	3,098,769

*We use patient-month here for consistency within the table, but this is the same as member-month.

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 show that when assessing reliability at the plan level, that the measure has good to very good reliability. ICC norms are: excellent reliability (≥0.90); very good reliability (≥0.85); good reliability (≥70); low reliability (<0.70).[1-3]

2b1. VALIDITY TESTING

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

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

☒ **Performance measure score**

☒ **Empirical validity testing**

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

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

Construct validity: For empirical validity testing we assessed construct validity in our primary analysis. We used CA Medicaid data and assessed the relationship between plan performance on the measure and plan performance on related and unrelated HEDIS measures. Related HEDIS measures included: Medication management for asthma; and vaccines in children. Medication management measures for asthma are the most closely related to asthma ED use, but are not reported separately for pediatric patients in HEDIS, so we would not expect perfect correlation. Therefore, we also looked at HEDIS pediatric vaccine status, which we would expect to be higher in health plans focusing on pediatric quality, including potentially pediatric asthma. Our unrelated measures included ACE monitoring, Cervical CA screening, and imaging for low back pain. We looked at these unrelated adult quality measures, as we would expect there to be a lower correlation with these measures (to the extent that plans might focus specifically on individual measures rather than focusing on all measures equally). We chose these unrelated adult quality measures based on data availability and focusing on measures that were less likely to be applicable to pediatric populations (e.g., ACE monitoring, Cervical CA screening).

We used publicly available data on HEDIS measures in CA plans, matching the plans by name in CA Medicaid data and in the publicly available data. We estimated plan performance using predicted performance for the plan-level random effect in the risk adjustment models, and then transformed that into a Z-score. We then used pairwise correlation coefficients to compare Z-scores to plan performance on HEDIS measures (reported as percentages). We were only able to assess these relationships for CA plans, as we did not have access to MA plan names in our dataset and so could not assess HEDIS performance for those plans. Because not all plans had publicly available data on all HEDIS measures, we only did this assessment for those with available HEDIS data (48 of 103 health plans).

Predictive validity (secondary analysis): In addition, we demonstrated predictive validity at the clinic level assessed in VT data, assessing whether a statewide QI learning collaborative focused on improving asthma care and management within primary care helped decrease asthma-related ED utilization. For this analysis, we used a difference in differences approach for practices in a learning collaborative focused on pediatric asthma compared to similar practices not in a learning collaborative in VT. The results of this work are in-press at *Pediatrics*. Citation: Harder V., et al. "Statewide Asthma Learning Collaborative Participation and Asthma-related Emergency Department Use". Pediatrics *In Press*.

Definition of numerator event: During Methods Panel Review, reviewers noted a concern regarding including in the definition of an asthma-related visit, visits with diagnosis of asthma in the second diagnostic spot. The concern was that a secondary diagnosis of asthma may be unrelated to the reason for ER visit/admission, and therefore the validity of the measure has not been established.

- **Our response was as follows: We addressed this concern in a peer-reviewed paper using this measure. We included claims with a second diagnosis of asthma because the primary diagnosis**

was often a related symptom (e.g., fever, wheezing) or a known asthma trigger (e.g., upper respiratory tract infection, pneumonia, influenza). We checked this assumption by tabulating the primary diagnoses for all the ED visits that were included with asthma in the second diagnostic spot, to confirm this. In addition, we performed sensitivity analyses to assess whether the primary findings of the paper changed if we only included ED visits with a primary diagnosis of asthma. The paper assessed the relationship between pediatric asthma ED utilization and diagnoses of anxiety and/or depression. We present the work from that paper below. (citation: Bardach et al. Depression, Anxiety, and Emergency Department Use for Asthma. *Pediatrics* 2019).

- **The sensitivity analysis demonstrated two main findings**
  1. **Relationships were similar using both definitions (primary analysis: ED visits with asthma in 1st or 2nd spot (Table 2), and sensitivity analysis: ED visits with asthma diagnosis in 1st spot only (eTable8 below);**
  2. **Dropping visits with asthma diagnoses in the second spot led to a loss of almost half of the numerator events.**

The first finding supports the validity of inclusion of the visits using asthma in the 1st of 2nd spot. The second finding supports the decision to keep the more liberal definition, in order to avoid losing half the events, which would limit the utility of the measure to statistically differentiate between health plans.

Methods:

Sensitivity analysis: "Because numerator events were identified using asthma as the first or second diagnosis, we separately re-ran the analyses only using ED visits with asthma as a first diagnosis in the numerator."

Results:

PRIMARY ANALYSIS

Table 2. Asthma-related ED visits rate per 100 child-years by patient characteristics

Mental health conditions	Asthma-related ED visits/100 child-years, Rate (unadjusted, 95% Confidence Interval)	Relative Rate of Asthma-related ED visits/100 child-years (adjusted, 95% Confidence Interval)	*P*-value*
No anxiety or depression	15.2 (14.1-16.3)	Reference	Reference
Anxiety only	18.6 (16.6-20.6)	1.22 (1.10-1.35)	< 0.001
Depression only	24.8 (20.7-28.8)	1.43 (1.23-1.62)	< 0.001
Anxiety and depression	30.5 (27.5-33.5)	1.80 (1.60-2.00)	< 0.001

* P-values for the multivariate analysis. Adjusted model included adjustment for age category, gender, insurance type, and chronic disease status, measured using the PMCA- Pediatric Medical Complexity Algorithm, excluding from the algorithm diagnoses of asthma, anxiety, and depression.

SENSITIVITY ANALYSIS

eTable 8. Asthma-related ED visits rate per 100 child-years, defining Asthma-related ED visits using the primary diagnosis only

Mental health conditions	Asthma-related ED visits/100 child-years, Adjusted Rate	95% CI Lower Limit	95% CI Upper Limit	P-value*
No anxiety or depression	7.7	7.2	8.3	Reference

Mental health conditions	Asthma-related ED visits/100 child-years, Adjusted Rate	95% CI Lower Limit	95% CI Upper Limit	P-value*
Anxiety only	8.5	7.3	9.7	0.203
Depression only	9.5	7.4	11.6	0.053
Anxiety and depression	11.0	9.3	12.7	<0.001

* p-value for results of multivariate comparison testing model, adjusted for age category, gender, insurance type, and chronic disease status, measured using the PMCA- Pediatric Medical Complexity Algorithm, excluding from the algorithm diagnoses of asthma, anxiety, and depression. CI: Confidence Interval.

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

Construct Validity (primary analysis):

Measure	Unrelated or related	Correlation coefficient	p-value	Number of plans included
MMA50	Related	0.12	0.42	48
MMA75	Related	0.13	0.37	48
Child Vaccines	Related	0.33	0.02	48
ACE Monitoring	Unrelated	0.05	0.75	48
Cervical Cancer Screening	Unrelated	0.04	0.79	48
Low Back Pain Imaging	Unrelated	0.05	0.74	47

MMA50: Medication Management for People 5-65 years with Asthma-Medication Compliance 50%

MMA75: Medication Management for People 5-65 years with Asthma-Medication Compliance 75%

Child Vaccines: Percent of children 2 years of age who were up to date on vaccines on their 2nd birthday.

ACE monitoring: Annual Monitoring for Patients on Persistent Medications-ACE Inhibitors or ARBs

Low Back Pain Imaging: Primary diagnosis of low back pain without an imaging study (plain X-ray, MRI or CT scan) within 28 days of the diagnosis.

Predictive Validity (secondary analysis, since not at the plan level):

In our analysis of VT practices, we found that there was an overall difference in difference for learning collaborative participating practices (n=20) compared to non-participating practices (n=15) between 2017 and 2014 (improved visit rates by 7.62 ED visits per 100 child-years) is statistically significant.

Table: Comparison of participating and control practice mean asthma-related emergency department visit rates from baseline and post-collaborative years (main difference-in-differences analysis)

Main Analysis	Asthma-related ED visits per 100 child-years	*P*-value	95% Confidence Interval: Lower	95% Confidence Interval: Upper
Control 2014	16.97	*	11.99	21.95
Participating 2014	15.29	*	11.62	18.96
Control 2017	18.80	*	12.89	24.71

Main Analysis	Asthma-related ED visits per 100 child-years	*P*-value	95% Confidence Interval: Lower	95% Confidence Interval: Upper
Participating 2017	9.50	*	7.07	11.93
Difference-in-Differences	*-7.62*	0.002	-13.45	-1.78

Notes: Marginal rates are adjusted for patient (age, sex, and insurance) and practice (specialty, organizational structure, size, geographic region, and federally qualified/rural health center) characteristics and also for patients clustering within practices; SE: standard errors are estimated using the delta method; Coefficients for relative rates are in Supplemental Table 2.

*cell intentionally left blank

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

Construct validity: This analysis supports construct validity, demonstrating greater correlation between the proposed measure of Pediatric ED Use and related measures than between Pediatric ED use and unrelated measures. While the correlations with the related measures are not strong correlations, we would not expect them to be strongly correlated, due to the multiple factors that go into Pediatric Asthma ED Use, and the fact that HEDIS measures of asthma medication use include adults as well as pediatric patients, and that immunizations for children <2 years old require different efforts than optimizing care for pediatric asthma patients. However, the correlation with the child vaccination measure is the strongest, at 0.33 and is statistically significant. The effect sizes for all the related measures in the Table are larger than the unrelated adult measures, indicating the construct validity of Pediatric Asthma ED Use.

Predictive validity: The analysis supports predictive validity, demonstrating that the measure is responsive to a QI initiative. Though this is a clinic-level analysis, a QI learning collaborative is a reasonable approach for a health plan to consider supporting across clinics as an approach to improving rates on this measure of Pediatric ED Use.

Definition of the numerator event: The analyses presented support the definition including as numerator events visits with asthma in the first or second diagnostic spot. The analyses provide evidence of the validity of the definition as well as the increased usability associated with the broader definition, in order to support adequate sample size for statistically meaningful differences between health plans.

2b2. EXCLUSIONS ANALYSIS

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

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

For exclusion of diagnoses, we assessed the percent of patients excluded based on the exclusion criteria (cystic fibrosis and emphysema). These diagnoses were chosen based on recommendations from a national advisory board convened as part of the measure development work in PQMP Phase 1. These exclusions led to very few patients being dropped (<2%).

Requiring 12 months of continuous enrollment is a common approach for measures of outcomes or utilization. However, we were concerned that: 1) requiring 12 months of continuous enrollment excludes a large number of patients and in particular, a more vulnerable patient population due to social risk factors, who would potentially benefit substantially from quality improvement efforts.

In addition, we were concerned that: 2) this continuous enrollment criteria incentivizes health plans to make it more difficult for patients to maintain enrollment, thereby providing a mechanism to exclude more vulnerable patients.

However, we also understand the concern from health plans that requiring less than 12 months of continuous enrollment would allow some patients to be attributed to a health plan that had not cared for the patient for very long and thus should not be held accountable for outcomes or utilization of these patients.

In order to test the Hypothesis 1, we assessed the percentage of patients who would be excluded if we held to the criteria of at least 12 months of continuous enrollment. We then assessed stratified rates for patients according to the number of months they were continuously enrolled.

Hypothesis 2 is not testable with our datasets.

In order to test Hypothesis 3, we assessed the patient-level mean number of months of continuous enrollment for each payer, after excluding patients with less than 3 months of continuous enrollment.

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

For exclusion of diagnoses: <1% of the otherwise eligible population in the MA dataset was excluded using these diagnoses.

For continuous enrollment criterion, using MA data:

Hypothesis 1:

Number of months of continuous enrollment	Rate of asthma ED visits per 100 child years*	Percent of otherwise eligible patients
<12 months continuous enrollment	18.5	51.2%
12 months continuous enrollment	16.3	48.8%

*Statistically significant difference by ttest at p=0.007

Hypothesis 3: The mean number of months of eligibility were similar across payers with at least 50 patients in MA (n=19 payers). The grand mean for these plans was 8.4 (95% CI 7.7-9.1) and SD 1.4. All these payers had a mean between 7.1-9.3 except for one plan outlier with a mean of 3.1 months of eligibility across 85 pediatric patients with asthma.

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

The number of exclusions by diagnosis is too small to make a difference through distortion of performance results.

The requirement of continuous enrollment of at least 3 months is preferable to requiring continuous enrollment of 12 months, based on retention of a much larger number of patients, and based on similar levels of continuous enrollment across plans with at least 50 patients.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

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

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

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with *6* risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

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

Risk model specifications:

Method: We used a negative binomial regression model, to account for the dispersed nature of the outcome.

Risk factors included patient and zip-code level factors. Definitions are listed in the Table below. Data for age, gender, and chronic condition indicator came directly from the claims data. Data for the zip code level variables came from the appropriate year of data from the American Community Survey results.

Variable	Definition	Source
Age	Age of member in years at first claim	Claims
Gender	Member gender at first claim in year. Options: male/female	Claims
Chronic condition indicator	Based on the Pediatric Medical Complexity Algorithm (Simon et al.). Three level categorical variable. No chronic disease, or non-chronic; non-complex chronic; complex chronic disease. Uses published ICD9 and ICD10 definitions. Assessed at the beginning of the measurement year.	Claims
% households below the poverty level	Using the patient 5 digit zip code, linked to data from the American Community Survey	American Community Survey
% population with less than high school education	Using the patient 5 digit zip code, linked to data from the American Community Survey	American Community Survey
% male unemployment for 25 to 60 year olds	Using the patient 5 digit zip code, linked to data from the American Community Survey	American Community Survey

We analyzed data from MA using STATA 16.0

Coefficients are as follows:

Variable	Coefficient
Age	-0.0141999
Female	0.0792704
Non-complex chronic condition	0.156758
Complex chronic condition	0.4730546
% households below the poverty level	0.021039
% population with less than high school education	0.0124768
% male unemployment for 25 to 60 year olds	-0.009898

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

Not applicable

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

Factors included in the baseline risk model are age, gender, and medical comorbidity status, as captured by the Pediatric Medical Complexity Algorithm, which is used to categorize patients into 3 categories by using International Classification of Diseases (ICD) codes: patients with no chronic conditions, noncomplex chronic conditions, or complex chronic conditions.[4] These factors were chosen a priori, based on the contributions of these characteristics to asthma severity.

Factors included for the baseline+socio-economic status (SES) model were: 5 digit zip-code level variables of social risk factors from the American Community survey, averaged over 5 years. The variables initially tested were: percent of households below the poverty level, percent of adults over 25 with less than a high school education, percent male unemployment for 25 to 60 year olds. These particular variables from the ACS have been validated as measures of SES using factor analysis and found to be associated with increased readmissions as well as direct measures of allostatic load or physiological stress.[5,6] Since race/ethnicity are not direct measures of social risk, it is not recommended per NQF or ASPE guidance to use them for social risk factor adjustment.

Ordering: We used a backward selection process, eliminating variables with p>0.10, starting with all factors. All variables were retained.

Note regarding offset term for exposure: The model does not include an offset term for the exposure (=the number of patient-year asthma per group). The underlying data structure for the analytic dataset (for which we have publicly available SAS code and a visual flow diagram) is a person-month structure. The use of an offset term would be appropriate if the data structure was data at the payer level, with the outcome being the number of admissions for the payer for the month, in which case one would want to know the denominator-the number of patient-years of asthma for the payer (since 72 admissions out of 72 patients in a month is quite different from 72 admissions out of 7200 patients). However, because the data structure for performance calculation is in the person-month level, we do not need an offset term for the modeling.

We identified 83,577 pediatric patients with asthma in MA, in 26 health plans. State-wide, there were 18.4 asthma-related ED visits/100 child-years. There were 2 health plans that moved to a higher performance ranking with social risk factor adjustment, compared to baseline risk-adjusted performance. No health plans moved to a lower performance ranking with social risk factor adjustment. Health plans whose performance improved with social risk factor adjustment had patients living in zip codes with higher poverty, lower incomes, lower educational achievement, and more unemployment (Table).

Table . Aggregate social risk measures for health plans with changed performance ranking after social risk factor adjustment

Health plans	Median household income	% pop with <high school education	% households below federal poverty level	% unemployed
All health plans (n=26)	$74,426	5.8%	12.1%	5.2%

Health plans	Median household income	% pop with <high school education	% households below federal poverty level	% unemployed
Health plans moved to **higher** performance ranking (n=3)	$60,217	7.8%	15.8%	5.9%

These analyses demonstrate that SES factors seem to drive performance, generally driving a change in performance for practices with lower SES patient populations.

Based on these analyses, we suggest including SES risk factors in comparative performance measurement.

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

☒ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

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

See above in 2b3.1.1.

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

See above in 2b3.3a.

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

We assessed the adequacy of the statistical model using the R-squared test, which quantifies the amount of variation in the outcome explained by the variables included in the model.

We calculated R-squared on plan-level performance assessment, to assess the predictive power for the model. To conduct this analysis, we used a plan-level dataset in each state, calculating the performance for the plan and the mean value for each risk adjustment variable across all member-months for members attributed to the plan. We performed linear regression, using the performance of the plan as the outcome, and using as predictors the mean values of the risk-adjustment variables across the member-months for each plan.

Note: The c-statistic is not appropriate for this model since the measure is a rate, not a proportion.

We assessed risk calibration by calculating the expected and predicted mean rates across deciles of patients, calculating a correlation coefficient using the pwcorr command in STATA (pairwise correlation coefficient), and creating visual plot to examine the plot of predicted vs. actual across deciles.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b3.9

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

R-squared values:

MA data: 0.56

CA data: 0.13

The interpretation is that using the MA data, the model explains 56% of the variance in the outcome (56%) and 13% of the variance in the outcome using the CA dataset (13%).

This indicates a stronger prediction model in MA than in CA, however, taken together with all the evidence presented in this attachment (validity, reliability, model calibration below, conceptual model for risk adjustment factors based on available literature), we chose to propose this model for the measure, given its performance on other testing.

During the Methods panel review, reviewers were concerned regarding the low R-squared values for the CA plans (13%), which is very different value from the R-squared value for the MA plans (56%). It was hypothesized that the model is a better fit for APCD than Medicaid only and that the Medicaid only population may have less variation in the community risk factors (% below poverty, education, unemployment) which may make the models less predictive.

In response, we assessed the variation in community risk factors in CA and MA, to test this hypothesis. We found the following:

Metric	MA mean	MA SE	CA mean	CA SE	Ratio of CA SE/MA SE
Median household income	66269	34.61233	54368.44	11.01509	0.318
% housing public assistance	17.01975	0.015342	5.908931	0.0021989	0.143
% households below FPL	14.54965	0.012076	17.40438	0.0054135	0.448
Number of households	10124.46	6.602852	14183.93	3.517135	0.533
% female headed household	15.71511	0.009999	17.55645	0.0032822	0.328
% unemployed	8.717494	0.004627	6.450902	0.0011429	0.247

***SE: standard error**

- **The ratio of CA standard error to MA standard error is consistently less than one, supporting the hypothesis that the Medicaid population in CA has less variation in community risk factors compared to the MA population. These findings support the hypothesis from the reviewer that the difference between the two states in their R-squared values is due to less variation across Medicaid members in CA compared to MA. This explains why the R squared values are dissimilar, with less of the variation in asthma ED visits explained by the risk-adjustment variables in CA compared to variation in asthma ED visits explained by risk adjustment variables in MA. These results should be taken in context with the rest of the validity data presented. While it would be nice to see the same R-squared across both states, it is not a fatal flaw in the measure, since R-squared is only one metric for judging validity.**
- **The rationale behind testing the measure in the CA dataset is to assess whether the model is overly specified for MA data. In order to test whether we had over-specified the risk adjustment model using MA data, we built the model afresh using CA data and the same backwards selection process. We started out with the same set of patient variables as in the base model (avg age, gender, chronic condition indicators) and the available community social determinants of health risk factors in the CA data (Median household income, % housing public assistance, % households below FPL, number of households, % female headed household,% unemployed). While the R-squared value went as high as 0.20 while including almost all risk factors in CA, the R-squared for the equivalent model in MA also improved similarly. This supports the idea that the difference in R-squared is not due to over-specification in the models, but rather that it reflects a more homogenous population in CA Medicaid alone vs. the members in the APCD data.**
- **Of note, we decided not to use the expanded set of variables from the CA backward selection process, since the original risk adjustment model reflected an evidence-based set of variables (**percent of households below the poverty level, percent of adults over 25 with less than a high school education, percent male unemployment for 25 to 60 year olds). These specific variables have been validated as measures of SES using factor analysis and found to be associated with increased

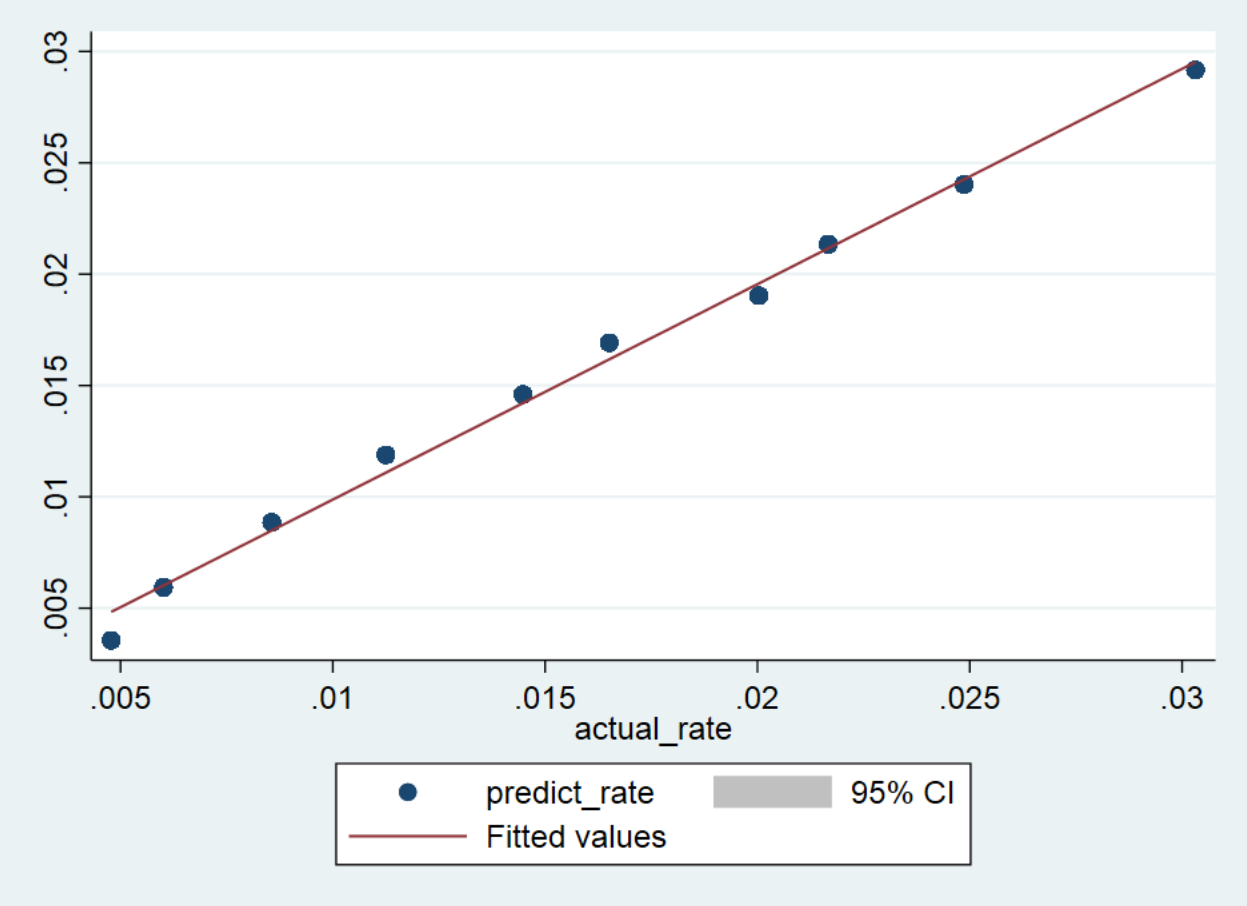
readmissions as well as direct measures of allostatic load or physiological stress (Martsolf, 2016 and Bird, 2010).

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

MASSACHUSETTS APCD DATA:

R=0.99 (correlation coefficient for predicted vs actual rates by decile of predicted)

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

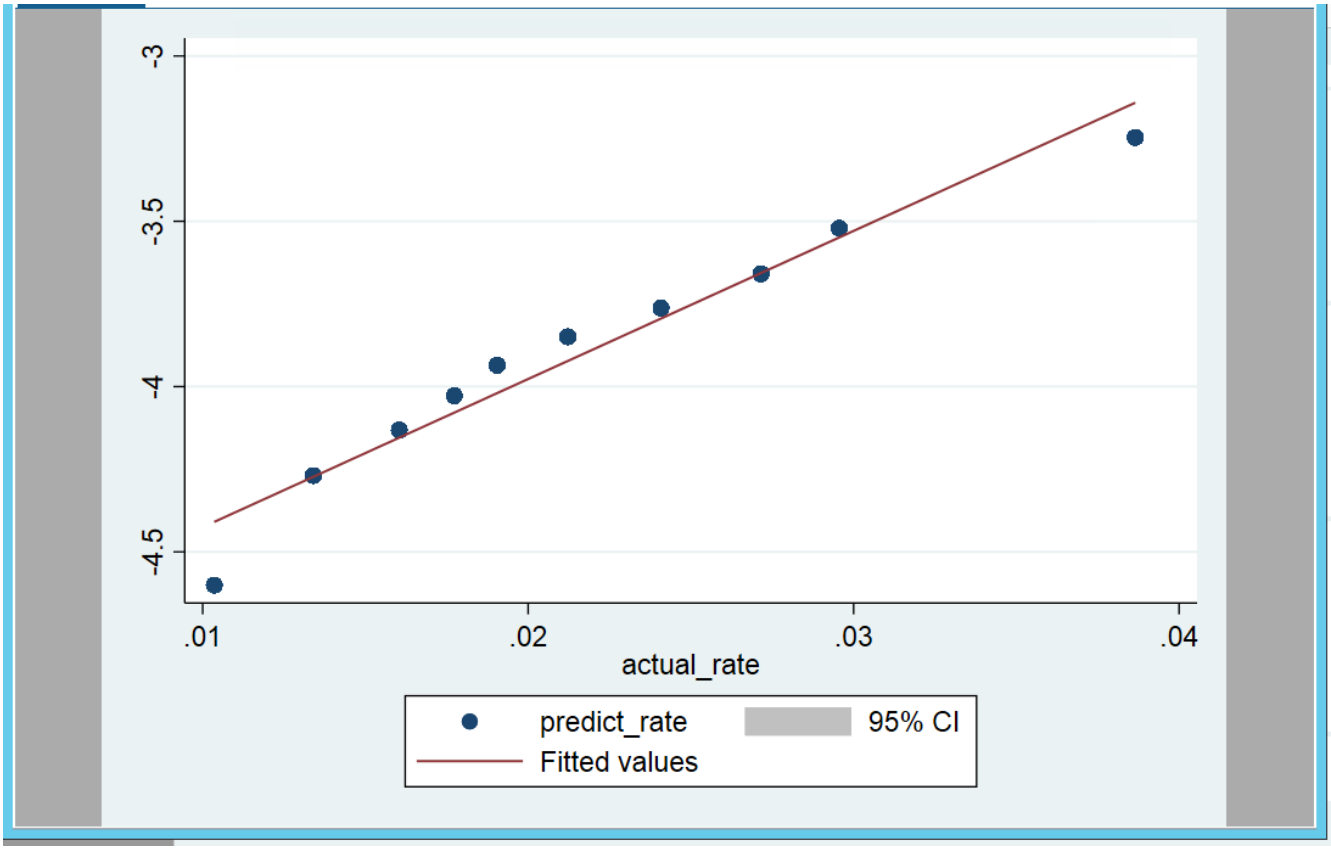


CALIFORNIA MEDICAID DATA

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

R=0.98 (correlation coefficient for predicted vs actual rates by decile of predicted)

Statistical Risk Model Calibration – Risk decile plots or calibration curves:



Note: the predicted and actual_rate values in the graph are not transformed. Transformed values are reported below in 2b3.9.

2b3.9. Results of Risk Stratification Analysis:

MASSACHUSETTS

Decile	Predicted #ED visits/100 child-years	Actual #ED visits/100 child-years
1	5.7	4.3
2	7.2	7.1
3	10.3	10.6
4	13.5	14.3
5	17.4	17.5
6	19.8	20.3
7	24.0	22.8
8	26.0	25.6
9	29.8	28.8
10	36.4	35.0

CALIFORNIA

Decile	Predicted #ED visits/100 child-years	Actual #ED visits/100 child-years
1	12.1	12.4
2	16.8	16.1
3	19.3	19.2
4	21.4	21.3
5	23.5	22.9
6	25.6	25.5
7	27.9	28.9
8	30.9	32.6
9	35.5	35.5
10	46.7	46.4

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

Our results show that the model adequately controls for differences in patient characteristics at the top and the bottom of risk strata groups.

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

Development of the model was conducted in MA data. We repeated analyses in CA data.

CA assessment is noted in each section above and are as follows:

Reliability: ICC in CA (Section 2a2.3) = 0.86 (0.79-0.90)

Discrimination: R-Squared in CA (Section 2b3.6) =0.13

Calibration: Correlation coefficient between predicted and actual in CA (Section 2b3.8): 0.98

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

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

We used standard z-score methodology to identify high, medium and low performers, based on CMS scoring for quality measures. We first fit a mixed effects negative binomial regression model with random effects for payer and fixed effects as noted above. We then generate the predicted effects and standard errors for each payer, in a post-estimation command.

We then calculated the Z-statistic for each plan.

Plans with a Z-statistic>1.96 were considered poor performing outliers, those with <-1.96 were considered high performing outliers and those in between were considered no different from average.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g.,

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

MASSACHUSETTS APCD DATA

outlier	Freq.	Percent	Rate Asthma ED visits
High performing	6	20.69	9.0
No diff from average	18	62.07	11.7
Low performing	5	17.24	24.2
Total	29	100.00	

CA MEDICAID DATA

outlier	Freq.	Percent
High Performing	16	15.38
No different from Average	55	53.85
Low performing	32	30.77
Total	103	100.00

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

There is meaningful variation across plans, with 40%-45% of plans identified as high or low performing (Section 2b4.2), and with clinically meaningful differences in performance (Section 2b3.9).

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

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

The measure has only one set of specifications.

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

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

NA

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

NA

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

what are the norms for the test conducted)

NA

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

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

Data was complete for age, sex, and chronic condition indicator for all patients.

Data on social risk factors was only missing for 6.6% of patients.

The level of missingness differed across plans (see Table in 2b6.2 below)

We dropped the plans with at least 40% missing to assess whether it changed results substantially for ICC, or for number of outliers identified.

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

Table. Distribution of missing data across plans in MA

Plan ID	Percent observations with *missing* social risk factor data	Total patients in plan
11715	0%	15
12226	0%	12
7397	0%	3
290	0%	3
3156	1%	17,776
3735	1%	13,447
3505	1%	7,762
4962	1%	6,670
11541	1%	178
296	2%	2,162
301	3%	4,930
8026	6%	564
10632	7%	1,377
300	9%	10,870
8647	10%	4,220
302	11%	942
7041	13%	45
10440	18%	53
312	20%	503
10441	20%	335
291	24%	13,257
11474	26%	196
10353	**47%**	**40**

Plan ID	Percent observations with *missing* social risk factor data	Total patients in plan
10444	**50%**	**228**
11939	**100%**	**85**
11943	**100%**	**9**
11936	**100%**	**9**

For MA, the overall rate of events statewide did not change with dropping of payers with >40% of patients with social factors missing, likely because there were so few patients in those plans (total n=371). The number of outliers identified was the same and the performance for each outlier group was the same.

We assessed missingness in CA Medicaid data as well:

Data was complete for age, sex, and chronic condition indicator for all patients.

Data on social risk factors was missing for 0.53%-0.58% of patients.

The level of missingness ranged across plans from a high of 3.31% to a low of 0%.

Due to the low level of missingness, we did not conduct further sensitivity analyses.

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

Our interpretation of our findings is that the effect of dropping the 6.7% of patients without SES social risk factors was not substantial. Though it excludes the 3 plans with 100% missing data from measurement, those plans had very few eligible patients and so any bias introduced would be minimal.

Our interpretation of this analysis is that the level of missingness in CA is not substantial.

Overall, though there is some missing data on social factors, this level of missingness is minimal and does not introduce a substantial problem in the use of this measure for health plan performance assessment.

References

1. Lyratzopoulos G, Elliott MN, Barbiere JM, et al. How can health care organizations be reliably compared?: Lessons from a national survey of patient experience. *Med Care.* 2011;49(8):724-733.
2. Merlo J, Chaix B, Ohlsson H, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epidemiol Community Health.* 2006;60(4):290-297.
3. Merlo J, Yang M, Chaix B, Lynch J, Rastam L. A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *J Epidemiol Community Health.* 2005;59(9):729-736.
4. Simon TD, Haaland W, Hawley K, Lambka K, Mangione-Smith R. Development and Validation of the Pediatric Medical Complexity Algorithm (PMCA) Version 3.0. *Academic Pediatrics.* 2018;18(5):577-580.
5. Martsolf GR, Barrett ML, Weiss AJ, et al. Impact of Race/Ethnicity and Socioeconomic Status on Risk-Adjusted Readmission Rates: Implications for the Hospital Readmissions Reduction Program. *Inquiry : A Journal of Medical Care Organization, Provision and Financing.* 2016;53:0046958016667596.

6. Bird CE, Seeman T, Escarce JJ, et al. Neighbourhood socioeconomic status and biological 'wear and tear' in a nationally representative sample of US adults. *J Epidemiol Community Health.* 2010;64(10):860-865.

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

Data on social determinants of health were missing for some patients. Please see Testing Attachment for results of missingness analysis and implications.

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)
Public Health/Disease Surveillance Regulatory and Accreditation Programs Quality Improvement (external benchmarking to organizations) Quality Improvement (Internal to the specific organization) Use Unknown	*

*cell intentionally left blank

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

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

NA

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

This is a new measure, and since not yet endorsed it is not currently in use. It is publicly available to all, with technical specifications and SAS code posted online for public use.

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

Dissemination and measure uptake are key goals of the Pediatric Quality Measurement Program. AHRQ provides leadership in developing and disseminating materials to facilitate uptake across health plans and

accountability programs. Specifically, AHRQ is working with LNM consulting to create a toolkit for this measure for dissemination and to encourage uptake by health plans. Judith Shaw, a co-investigator on the team, plays a leadership role in the National Improvement Partnership Network (NIPN) and will share the measure and measure toolkit with that group. Endorsement by the NQF will facilitate additional use in accountability programs, and may lead to potential inclusion in the Child Core Set for Medicaid plan measurement.

Timeline: We are meeting with a number of state Medicaid Medical Directors in the next few months to share the data on importance, validity, feasibility, and usability, based on the work we have conducted under the PQMP. We anticipate that there will be interest in implementation in at least one state, with potential use of the measure within 18-24 months. We will likely start with Medicaid managed care plans for internal reporting and then move towards public reporting, depending on interest in the state Medicaid offices.

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

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

The IMPLEMENT team, led by Dr. Michael Cabana, and funded by PQMP to further develop and refine this measure through feasibility and usability work, convened a CA asthma QI collaborative for CA-based pediatric primary care practices. There were 9 participating practices from northern, central and southern California. As part of the collaborative, there was an in-person all-day kick off meeting in March 2018, which included education on asthma management, evidence-based practices, and QI methodology. In addition, performance results on the proposed measure of Pediatric Asthma ED Use for Medicaid patients from the prior year were presented to participating practices. Results of performance were reviewed at the kick off meeting. Each site received a report showing their own performance and the overall group performance, and were shown the overall state performance, county performance within the state, and the de-identified performance of the other clinics. Measurement experts and asthma experts on the team then led a moderated group discussion, reviewing the validity of the results and interpretation of the data and discussion of how to approach improving performance on the measure.

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

See above.

In addition, we are waiting to receive 2019 data from CA Medicaid office in order to report on post-collaborative performance, as the collaborative ran for 18 months. We will provide the same report for clinics as they received in the initial meeting.

To inform ongoing improvement work during the collaborative, practices focused on asthma processes of care in their clinics, assessed through chart review, as the claims-based ED visit data has a lag period to availability. The process measure included the following: 1) The percentage of patients with asthma severity documented as intermittent or persistent (mild, moderate, or severe); 2) The percentage of patients prescribed inhaled corticosteroids or other control medications if asthma severity is persistent; 3) The percentage of patients with asthma control assessed with a validated tool; 4) The percentage of patients with an asthma action plan initiated, reviewed or updated as needed within the last 12 months; 5) The percentage of patients with at least one planned asthma visit every months; 6) The percentage of patients assessed for tobacco use/exposure; 7) The percentage of patients and their caregivers educated about their asthma; 8) The percentage of patients and their caregivers instructed on how to use their asthma delivery device. Measures 1–4 were required and teams chose any two of the remaining five to improve at their practice.

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

Describe how feedback was obtained.

After the quality improvement collaborative was complete, IMPLEMENT team leadership conducted semi-structured qualitative interviews with physician and QI champions at each of the participating clinics.

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

Sites that found the data helpful said:

--it made them aware of the magnitude of the problem of asthma ED use for their clinic population in a way not previously possible.

--it motivated them to look more closely at processes of care

--it was a good way to start the collaborative, in order to frame the importance of the project

--Provided a motivation for participation in the collaborative, since performance for the clinic was a lot higher than the state average.

Suggested improvements included:

--include data on urgent care visits, since some clinics have an urgent care that manages asthma exacerbations most of the time.

--the data from one of the sites was difficult to get and they were not confident in the number of patients (they seemed very low, and the site knew there were more patients for the denominator)

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

NA

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

In response to the national advisory council members, we included social determinants of health variables into the risk adjustment model, following the NQF and ASPE guidance on considerations around data sources and rationale for inclusion vs. not including these variables. We used an evidence-based approach to including these variables (see testing document).

Improvement

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

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

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

Improvement on this measure was associated with participation in the Vermont state-level quality improvement collaborative, as presented in the Evidence attachment.

4b2. Unintended Consequences

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

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

None

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

None

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

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

Yes

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

0728 : Asthma Admission Rate (PDI 14)

1381 : Asthma Emergency Department Visits

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

1381 is no longer endorsed. Endorsement last updated 2014. Measure title: Asthma Emergency Department Visits. Steward: Alabama Medicaid Agency

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.

Regarding measure 0728: Full technical specifications are not available as this measure is being reviewed for maintenance of endorsement. However, the measure we propose focuses on a different types of utilization, ED use, rather than asthma hospitalizations. Measure 0728 is also intended for population level analysis at the regional or state level, which differs from the use case for the proposed measure, which is health plan use, generally in collaboration with primary care practices.

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.)

NA

Appendix

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

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

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Albert Einstein College of Medicine

Co.2 Point of Contact: Michael, Cabana, mcabana@montefiore.org

Co.3 Measure Developer if different from Measure Steward: University of California San Francisco

Co.4 Point of Contact: Naomi, Bardach, naomi.bardach@ucsf.edu

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.

Asthma Advisory Council:

Members of the asthma national advisory council met with the measure developer and steward twice annually to inform testing and refinement of the measure. They gave expert advice regarding the validity and usability of the measure regarding improving asthma care and outcomes.

Members were:

Barbara Yawn MD, MSc Adjunct Professor of Family and Community Health at the University of Minnesota

Elizabeth Cox MD, PhD Associate Professor of Pediatrics at the University of Wisconsin

Lisa Cicutto BSN, MSc, PhD Director of Community Outreach and Research at National Jewish Health the Director of the Clinical Science Graduate Program at the University of Colorado Denver

Joseph Zorc MD, MSCE Pediatric Emergency Medicine Physician and a Professor of Pediatrics at the University of Pennsylvania

Keith Robinson MD Assistant Professor of Pediatric Pulmonology, Vice Chair of Quality Improvement and Population Health, University of Vermont Children´s Hospital

Judith Shaw EdD, MPH, RN, FAAP Executive Director of Vermont Child Health Improvement Program (VCHIP), Professor of Pediatrics and Nursing, UVM Health

David Brousseau MD, MSProfessor of Pediatrics and Chief of the Section of Emergency Medicine at the Medical College of Wisconsin

Jernee Carter Parent of a child with asthma

Usability advisory council:

Members of the usability council met with measure developer and steward twice annually to inform testing and refinement of the measure. They gave expert advice regarding the usability of the measure for use in

primary care quality improvement efforts and quality improvement collaboratives, providing a wide range of stakeholder perspectives, including those of EQROs, health plans, quality improvement officers, and others.

Members were:

Virginia Moyer MD, MPH Vice President for Maintenance of Certification and Quality at the American Board of Pediatrics. She served as the first Chief Quality Officer for Medicine at Texas Children´s Hospital

Maria Britto MD, MPH Professor of Pediatrics and Founding Director of the Center for Innovation in Chronic Disease Care

Nora Wells MS Ed Executive Director of the National Office of Family Voices

Mary Fermazin MD, MPA Chief Medical Officer, Health Services Advisory Group, Inc. (HSAG)

Jim Glauber MD, MPH Chief Medical Officer at San Francisco Health Plan

Susan Fleischman MD Chief Medical Officer at Blue Shield Promise Health Plan

Barsam Kasravi MD, MBA, MPH Medical director in the area of Clinical Quality and Innovation at Blue Cross of California

Judith Shaw EdD, MPH, RN, FAAP Executive Director of Vermont Child Health Improvement Program (VCHIP), Professor of Pediatrics and Nursing, UVM Health

Irwin Charles MD, MPH Distinguished Professor of Pediatrics, Director of the Division of Adolescent Medicine and Adolescent Health at the University of California, San Francisco

Margaret Morris MA, CHCA Managed Care Senior Director on the Pediatric Value-Measurement Advisory Panel of the Washington DC-based Children's Hospital Association

Feasibility advisory council:

Members of the feasibility council met with measure developer and steward twice annually to inform testing and refinement of the measure. They gave expert advice regarding the use of administrative claims in performance measurement for this measure, risk adjustment model choices, including the choice of social determinants variables for risk adjustment, and methods for identifying performance outliers.

Members were:

Patrick Romano MD, MPH Professor of Medicine and Pediatrics at the University of California, Davis

Adams Dudley MD, MBA Professor of Medicine and Director of the University of California, San Francisco (UCSF) Center for Healthcare Value (CHV)

Susan Paulukonis MA, MPH Program Director at the California Rare Disease Surveillance Program at the Public Health Institute and the California Department of Public Health

Judith Shaw EdD, MPH, RN, FAAP Executive Director of Vermont Child Health Improvement Program (VCHIP), Professor of Pediatrics and Nursing, UVM Health

Chuck McCulloch PhD Professor and Head of the Division of Biostatistics and Vice Chair of the Department of Epidemiology and Biostatistics at the University of California, San Francisco

Valerie Harder PhD, MHS Director of Health Services Research at the Vermont Child Health Improvement Program (VCHIP)

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2018

Ad.3 Month and Year of most recent revision: 10, 2019

Ad.4 What is your frequency for review/update of this measure? Every 2 years

Ad.5 When is the next scheduled review/update for this measure? 10, 2021

Ad.6 Copyright statement: This measure is available in the public domain.

Ad.7 Disclaimers: none

Ad.8 Additional Information/Comments: none